Frontiers
in
Artificial
Intelligence
and
Applications

MODERN MANAGEMENT BASED ON BIG DATA IV

Proceedings of MMBD 2023

Edited by
Antonio J. Tallón-Ballesteros

IOS Press

MODERN MANAGEMENT BASED ON BIG DATA IV

The concept of Big Data has become increasingly familiar in recent years, and it is already an indispensible tool in the management of everything from supply chains and transport to health and education.

This book presents the proceedings of MMBD 2023, the 4th International Conference on Modern Management based on Big Data, held in Seoul, South Korea, from 1-4 August 2023. The 50 papers included here were selected from total of around 160 submissions after a rigorous review process. Papers delivered at the conference were divided into 3 main categories: Big Data, Modern Management, and a special session devoted to Big Data-driven manufacturing and service-industry supply-chain (SC) management, but in addition to these general topics, there were also a number of papers related to lifelong education. Topics covered in the book include innovation in online education management with big data; digital transformation in lifelong education; big data analysis in lifelong education management; green supply chain management; big data analytics in supply chains; policy and strategy for new energy and the environment; smart grid load and energy management; decision-making on sustainable transport policies; modern healthcare management; and social strategy to manage human relationships. Of particular interest are papers concerning big-data analysis and emerging applications.

Presenting innovative original ideas and methods, together with significant results, and supported by clear and rigorous reasoning and compelling new evidence, the book will be of interest to all those who use Big Data to support their management strategies.

9 781643 684109

ISBN 978-1-64368-410-9 (print)
ISBN 978-1-64368-411-6 (online)
ISSN 0922-6389 (print)
ISSN 1879-8314 (online)

MODERN MANAGEMENT BASED ON BIG DATA IV

Frontiers in Artificial Intelligence and Applications

The book series Frontiers in Artificial Intelligence and Applications (FAIA) covers all aspects of theoretical and applied Artificial Intelligence research in the form of monographs, selected doctoral dissertations, handbooks and proceedings volumes. The FAIA series contains several sub-series, including 'Information Modelling and Knowledge Bases' and 'Knowledge-Based Intelligent Engineering Systems'. It also includes the biennial European Conference on Artificial Intelligence (ECAI) proceedings volumes, and other EurAI (European Association for Artificial Intelligence, formerly ECCAI) sponsored publications. The series has become a highly visible platform for the publication and dissemination of original research in this field. Volumes are selected for inclusion by an international editorial board of well-known scholars in the field of AI. All contributions to the volumes in the series have been peer reviewed.
The FAIA series is indexed in ACM Digital Library; DBLP; EI Compendex; Google Scholar; Scopus; Web of Science: Conference Proceedings Citation Index – Science (CPCI-S) and Book Citation Index – Science (BKCI-S); Zentralblatt MATH.

Volume 370

Recently published in this series

Vol. 369. P. Hitzler, M.K. Sarker and A. Eberhart (Eds.), Compendium of Neurosymbolic Artificial Intelligence
Vol. 368. P. Lukowicz, S. Mayer, J. Koch, J. Shawe-Taylor and I. Tiddi (Eds.), HHAI 2023: Augmenting Human Intellect – Proceedings of the Second International Conference on Hybrid Human-Artificial Intelligence
Vol. 367. A.J. Tallón-Ballesteros and P. Santana-Morales (Eds.), Digitalization and Management Innovation – Proceedings of DMI 2022
Vol. 366. R. Hakli, P. Mäkelä and J. Seibt (Eds.), Social Robots in Social Institutions – Proceedings of Robophilosophy 2022
Vol. 365. L.C. Jain, V.E. Balas, Q. Wu and F. Shi (Eds.), Design Studies and Intelligence Engineering – Proceedings of DSIE 2022
Vol. 364. M. Tropmann-Frick, H. Jaakkola, B. Thalheim, Y. Kiyoki and N. Yoshida (EDS.), Information Modelling and Knowledge Bases XXXIV
Vol. 363. A.J. Tallón-Ballesteros (Ed.), Proceedings of CECNet 2022 – The 12nd International Conference on Electronics, Communications and Networks (CECNet 2022), 4–7 November 2022

ISSN 0922-6389 (print)
ISSN 1879-8314 (online)

Modern Management Based on Big Data IV

Proceedings of MMBD 2023

Edited by

Antonio J. Tallón-Ballesteros

University of Huelva, Huelva city, Spain

IOS Press

Amsterdam • Berlin • Washington, DC

ISBN 978-1-64368-410-9 (print)
ISBN 978-1-64368-411-6 (online)
doi: 10.3233/FAIA370

Publisher
IOS Press BV
Nieuwe Hemweg 6B
1013 BG Amsterdam
Netherlands
fax: +31 20 687 0019
e-mail: order@iospress.nl

For book sales in the USA and Canada:
IOS Press, Inc.
6751 Tepper Drive
Clifton, VA 20124
USA
Tel.: +1 703 830 6300
Fax: +1 703 830 2300
sales@iospress.com

Preface

Organized by Keimyung University, and co-organized by Beijing Wuzi University and the University of Messina (UniME), the 4th International Conference on Modern Management based on Big Data (MMBD2023) was held from 1-4 August 2023 in Seoul, South Korea. With MMBD2023, the conference series Modern Management based on Big Data (MMBD) completed its fourth edition.

The general topics of MMBD2023 fall into three main categories: 1) Big Data, 2) Modern Management and 3) a Special Session on big data-driven manufacturing and service industry supply chain (SC) management.

Apart from the general topics, MMBD2023 attracted a number of papers related to lifelong education. Topics at the conference included but were not limited to: innovation in online education management with big data; digital transformation in lifelong education; big data analysis in lifelong education management; green supply chain management; big data analytics in supply chains; policy and strategy for new energy and environment; smart grid load and energy management; decision making on sustainable transport policies; modern healthcare management; and social strategy to manage human relationships.

The most popular topics, both at the conference and in this book, concern big data analysis and emerging applications.

All of the papers were exhaustively reviewed by program committee members and peer-reviewers, who took into account the breadth and depth of the research topics that fall under the scope of MMBD. The 50 most promising and FAIA mainstream-relevant contributions were selected from about 160 submissions for presentation and inclusion in this book, which presents innovative original ideas or results of general significance supported by clear and rigorous reasoning and compelling new evidence, as well as methods.

I would like to thank all the keynote and invited speakers, authors, program committee members and anonymous reviewers for their efforts in making MMBD a conference of the highest standard.

June 2023

Antonio J. Tallón-Ballesteros
University of Huelva (Spain)
Huelva city, Spain

This page intentionally left blank

About the Conference

Organized by Keimyung University and co-organized by Beijing Wuzi University and the University of Messina (UniME), the 4th International Conference on Modern Management based on Big Data (MMBD2023) was held from 1-4 August 2023 in Seoul, South Korea.

All papers were conscientiously reviewed by members of the technical program committee (TPC) and peer-reviewers. These reviewers bore in mind the breadth and depth of the research topics that fall under the scope of MMBD, and are listed on this link: http://www.academicconf.com/ReviewerList?pageIndex=1&pageSize=50&confName=mmbd2023

From a total of around 160 submissions, the 50 most promising MMBD and FAIA mainstream-relevant contributions are included in this book, which presents original ideas or results of general significance supported by clear reasoning, compelling evidence, and methods.

General Chair

Prof. Dr. Yoonmin Kim, Department of Economics and Finance, Keimyung University, South Korea; Vice Director at Asia Center & Vice Director for FTA Specialist Education Program, Keimyung University, South Korea

TPC Co-Chairs

Prof. Dr. Bo Lyu, Dean, Business School, Beijing Wuzi University, China
Prof. Dr. Yuan Sun, Secretary of CPC Committee and Deputy Dean of School of Business Administration, Zhejiang Gongshang University, China
Prof. Dr. Antonino Galletta, University of Messina, Messina, Italy

Technical Program Committee (TPC)

Dr. Luis-Alberto Casado-Aranda, Assistant Professor, Department of Marketing and Market Research, University of Granada, Spain
Dr. Omid M. Ardakani, Associate Professor, Parker College of Business, Georgia Southern University, USA
Dr. Fuad Aleskerov, Professor, National Research University Higher School of Economics (HSE University), Russian Academy of Sciences, Russia
Dr. Antonio J. Tallón-Ballesteros, Department of Electronic, Computer Systems and Automation Engineering, University of Huelva, Spain

This page intentionally left blank

Contents

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230165

Research on Forecasting Sales of Products Based on Spatiotemporal Graph Neural Network

Shao-yan ZUO[a], Hai-fu HUANG[a], Lin GAO[b,1], Hai-ying LI[c], Hai-bo WANG[c], Chen ZHANG[c], Zhang-sheng WU[c]
[a]*China Tobacco Guangxi Industrial Co., Ltd., Guangxi 530001, China*
[b]*Shandong Qingdao Tobacco Co. LTD, Shandong 256400, China*
[c]*Zhongke Zhidao Technology Co., LTD., Beijing 100190, China*

Abstract. Intelligent marketing and recommendation are a core business of commercial companies, and accurate prediction of sales is the premise and foundation for greater efficiency of smart marketing and recommendation. In order to predict product sales, deep neural network (DNN), convolutional neural network (CNN), time series analysis and other methods have been put forward, but most of which only focus on the temporal or spatial characteristics of data. According to modeling and analyzing sales of products, they are closely related to the spatial location and time of the corresponding merchants. The goal is to predict the sales of products accurately at a given time and place, we advance a hybrid model of CNN-LSTM to forecast sales. Firstly, a large-scale knowledge graph system based on merchants is constructed, which describes the sales data and the relevant interaction scenarios of the corresponding business, merchants and users through the data model of a graph, and add the spatial and data characteristics of the business data on the graph model to describe the temporal and spatial characteristics of the merchants. Based on the constructed business knowledge graph, graph convolutional neural network (GCN) is used to aggregate information and obtain spatial features. Correspondingly, long short-term memory (LSTM) is used to extract time features. Researchers combine the two characteristics to make the sales forecast. In this study, neural network and GCN-LSTM algorithm are respectively used to carry out experiments on two kinds of product regulations. The result shows that the sales predicted by hybrid model of GCN-LSTM is almost as equal as the actual sales. The average accuracy of the proposed model is 89%.

Keywords. Graph neural network, Sales prediction, Long-short term memory, Product, Big data

1. Introduction

Marketing and recommendation are the core of commercial companies as well as the premise prediction for accurate sales. In the past, commodity marketing and recommendation were characterized by a lot of man-power, material and financial resources due to a huge amount of business. At the same time, the repetitive and heavy

[1] Corresponding Author: Lin Gao, Shandong Qingdao Tobacco Co. LTD, Shandong 256400, China. Email: watermelongl@qq.com, 270377937@qq.com

business have encountered some difficulties in the market. Therefore, we propose an intelligent method to predict and recommend some products to meet the needs of merchants so as to predict sales accurately. Many researchers [1-6] devote themselves to it, for instance, LSTM network model was built which was convenient for data input with great accuracy in time series prediction. A cost-aversion biased combination prediction model was proposed on the basis of random forest, GBDT and XG-Boost algorithms, which could accurately predict sales. The time series model based on multi-population genetic algorithm proposed that showed a high prediction accuracy for commodities. Deep learning prediction models combined with pictures and other structured data are proposed to advance a more accurate sales prediction method. However, there are few systematic studies on the combination of Spatio-temporal characteristics of product marketing. Graph convolution network (GCN) and long short-term memory (LSTM) spatiotemporal network were used for rolling frequency prediction [7]. We construct a merchant knowledge map to represent the relationship among stores. And GCN-LSTM hybrid model is constructed to extract spatial features and temporal features, which will be combined with sales prediction. We raise this method that can obtain more ideal marketing forecast.

2. Overview of research

Forecasting commodity marketing is one of the most important businesses among commercial companies. Traditional forecasting methods inevitably have heavy data workloads and manual limitations in fetching, memory, calculation, etc. It is difficult to achieve a large-scale accurate calculation, which will directly affect the economic benefits of commercial companies and merchants. Therefore, artificial intelligence methods for accurate predictive marketing have been the key research direction of various companies.

Deep neural network has a powerful feature extraction [8], graph convolutional neural network (GCN) has shown a strong application performance as a deep learning representation algorithm, which not only express complex semantic relations, but also capture global graph information [9]. LSTM is widely used in the field of artificial intelligence, multiple LSTM units forming an LSTM network can be used to learn the characteristics of the temporal dimension of the input spatiotemporal data.

In order to predict sales of commodities accurately, different data types of commodities are used as input. In this paper, a new framework based on graph convolutional neural network (GCN) and long short-term memory (LSTM) neural network is proposed for the prediction of time and space sales. Specifically, the main work of this study is organized as follows:

(1) Perform feature fusion and process data, complement 0 for numerical data, and perform log1p function to obtain smooth data relatively, and class data is supplemented with null, encoded by Label-Encoder, and converted to obtain numerical features. In neural network A. kaiming-normal is used for weight initialization and other operations. After the fusion of features, the data is input into neural network B and passed through ReLU respectively. Batch normal layer normalization and dropout operation are processed to obtain the difference between the measured target value and the predicted value.

(2) The prediction method of graph neural network is proposed. Construct many store knowledge graphs in the business district, extract node and the edge of the attribute

to predict relationships between a node and an edge, after making a knowledge fusion, Graph convolution neural network algorithm is proposed. We establish knowledge graphs to input data, neighbors of each node are used by convolution operation, convolution results update the node, realize information aggregation, the nonlinear activation function is used for two-layer convolution to achieve the expected depth.

(3) A GCN-LSTM hybrid model is constructed for the prediction of product sales by graph convolutional neural network. LSTM is used to solve the problem that long-term dependent feature relations cannot be learned. The proposed model fuses GCN related spatial data and LSTM time data to create a new data, which makes full use of the location and the structure of shops, LSTM capture the characteristics of the dynamic sales change from several aspects of time correlation to achieve the ultimate for goods sales forecast, and improve the accuracy of commodity sales prediction.

(4) In order to express the accuracy about intelligent prediction of product sales, two kinds of products with different product regulations are carried out relevant visual experimental verification of sales prediction in different time periods. In contrast to actual sales, it is shown that prediction accuracy of GCN-LSTM model is 89% on the average.

3. Algorithm of Sales forecasting based on shallow neural network

With the progress of development of information technology, more and more methods can be used to deal with the prediction such as BP neural network, time series analysis prediction, prediction based on support vector machine, etc. Although these methods have good effects in some aspects, they still have many shortcomings. However, neural networks in deep learning can effectively solve the limitations of traditional methods, deal with the prediction tasks with high complexity, and the prediction results are closer to the real value. Therefore, this study proposes a product delivery algorithm based on shallow neural network, which is used to predict product sales and achieve a more accurate marketing forecast of products. The flow chart of product delivery algorithm based on shallow neural network proposed in this study is shown in Figure 1.

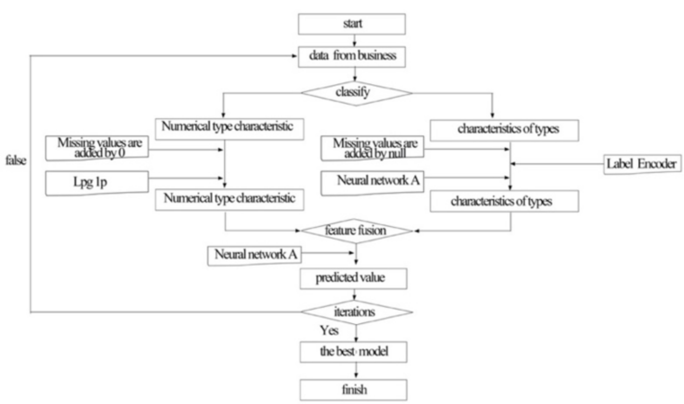


Figure 1. Flow chart of product delivery algorithm based on shallow neural network

Label Encoder transforms the categorical data into numerical data. If there is a label of a mobile phone brand, under which there are "Apple", "Huawei" and "Xiaomi", Label Encoder will be called to map "Apple", "Huawei" and "Xiaomi" respectively from 0, and the result is 0, 1, 2. The structure of neural network A is shown in Figure 2.

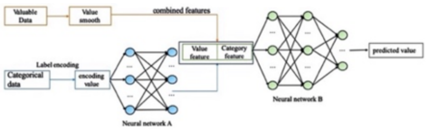


Figure 2. Neural network architecture diagram.

Figure 3. The knowledge graph of Store

Neural network A is a two-layer neural network, and Kaiming- normal which is a particularly robust initialization method that adapts to the nonlinear activation function is used for weight initialization. After the output of the first layer neural network, batch-normal function is used to normalize the output to the standard distribution shape, and ReLU is used to activate the output. After the output of the second layer neural network is inserted, the output of the second layer neural network is regarded as a set of category features generated by the neural network after learning. The smoothed numerical feature will be fused with the category feature. Since both the numerical feature and the neural network feature are vectors, this algorithm spliced the two feature vectors, as shown in Formula (1), where NFi represents the numerical feature, EFi represents the category, and CFi represents the fusion feature. NFi can be interpreted as $NFi^T = (NF_1, NF_2 \dots\dots, NF_n)$, $EFi^T = (EF_1, EF_2 \dots\dots, EF_m)$,whereas CFi is m plus n vector.

$$CF_i = [NF_i, EF_i] \tag{1}$$

After the features are fused, they are fed into neural network B. The neural network is a three-layer structure, and kaiming-normal is also used for weight initialization. In the output of the first layer, the ReLU is used to activate the dropout operation, and the output is sent to the batch-normal layer for normalization. Then, the output of the second layer is sent to the God network, and the output of the second layer is also performed a dropout operation. Finally, the output of the fully connected layer is connected and the output of the fully connected layer is used as the predicted value. The entire neural network is trained using MSE Loss, which measures the difference between the target value and the predicted value, as shown in Equation (2), where x is the predicted value and y is the true value.

$$loss(x_i, y_i) = (x_i - y_i)^2 \tag{2}$$

The neural network optimizer uses Adam optimizer with a learing rate of 0.0001, and 64 groups of data are input for algorithm training in each batch. We have tried another optimizer to test the results such as SGDM, SGD, Adagrad, etc. However, we find the Adam shows the best performance in the real application. For your references, we have listed the contrast in table 1.

Table 1. Performance analysis of different optimizers in deep learning model of commodity sales prediction

Optimizer	SGD	Adam	Adagrad	SGDM
Time/s	12.212	12.158	14.213	13.272
Loss	0.4171	0.3511	0.5355	0.4388
accuracy	0.8451	0.8964	0.8815	0.8317

4. Convolutional neural network algorithm

4.1 Figure

A graph is a data structure that describes the relationship between the one and the other one. It is usually composed of nodes and edges, where nodes represent entities, and edges between nodes indicate that there is relationship between two entities. A graph can be also expressed as $G = \{V, \mathrm{E}\}$, where V is the set of all nodes in the graph, E represents an edge set that describes the relationship between the nodes and the adjacent node. As Formula (3) shown, the graph has many nodes, $A = (a_{ij})_{n\times n}$,which represents the adjacency matrix of the graph.

$$a_{ij} = \begin{cases} 1\,, v_i \; and \; v_j \;\; are \; connected \\ 0, v_i \; and \; v_j \; aren't connected \end{cases} \tag{3}$$

If graph is an undirected graph, $a_{ij=}a_{ji}$, graph represent the degree matrix, and denotes the number of connected nodes. Assuming that each node in the graph has a feature, the features of all nodes in the graph are combined into a matrix.

4.2 Business area information atlas construction

The knowledge graph includes knowledge modeling, knowledge extraction, knowledge fusion, knowledge storage, knowledge reasoning, and knowledge application [10]. The “entity-relational-entity” is represented by SPO triplet, and the feature information comes from the attributes of nodes and edges. This form has a strong interpretability and can directly reflect the structural and semantic information of the atlas. Based on the store information, the spatial size data collected in this study are shown in Table 2, the knowledge graph constructed is shown in Figure 3, and the overall framework is shown in Figure 4.

Table 2. The capacity of Knowledge graph

Number of the store	Number of products	Node	Edge
4	2	10	9

In the process of building store knowledge graph, the data used are multi-source and heterogeneous. Structured data comes from the integration of research teams. Semi-structured data comes from relevant data of commodity delivery platform, which needs to be normalized by attributes. Unstructured data comes from fragmented text content, and knowledge processing needs to extract information from retail label data in the business area. After the data source is obtained, the knowledge is modeled. It is defined as Conceptual Domain - Entity Domain - Event Domain. The conceptual domain is the abstraction of the concrete entity, and the edge between the two entities is represented by the distance between the stores, the product regulation and the sales volume respectively. Entity domains are business-relevant instances, such as different store names, product specifications, and sales volumes in the graph. The event domain refers to the purchasing behaviors of customers. The events behind these behaviors are used as the precipitation of structured knowledge to enhance some static knowledge in the entity domain.

The third part is knowledge processing. For semi-structured or unstructured business data, the acquisition of knowledge will involve classification or extraction, such as how to define the relationship between each store entity. After knowledge acquisition, the structured data of knowledge is mapped to the architecture in knowledge modeling. Then do the entity chain refers to the normalization work, one is the entity link, two is the entity normalization or attribute normalization. After knowledge processing, triplet data is fused to make knowledge inference on complex semantic relations, and attributes and relations are predicted: two stores less than 3 kilometers can correspond; Shops belong to the area of the business district, and then predict the rules and sales of a product, the dragon (hard Ling-yun) belongs to the sales rules in the store, and its sales volume is 443 boxes.

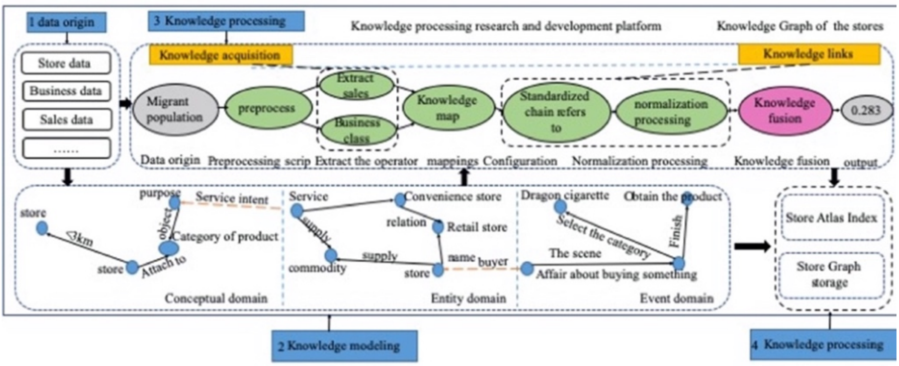


Figure 4. Overall framework of the knowledge graph

4.3 Overview of neural network

Convolutional neural network extracts data features through convolution. However, classical convolutional neural networks can only process Euclidean spatial data such as images and texts among others which are of translation invariance. In order to process non-Euclidean spatial data such as graph data, F. Scarselli have proposed the graph Neural network (GNN) [11] method. GNN can transform graph- structure data and then input them into various neural networks for training. Graph neural network can be classified [12] into Graph Convolution Network, Graph Attention Network, and Scalable Graph Network.

4.3.1 Figure Convolutional Neural network

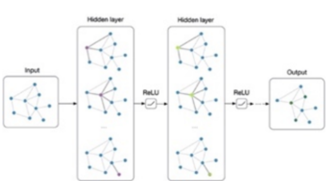


Figure 5. Training process diagram of GCN

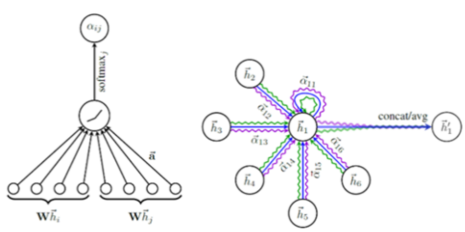


Figure 6. The learning process of attention network

Bruna [13] et al. proposed a method to combine graph and convolutional neural networks i.e. graph convolutional neural network (GCN) in 2013which aims to accomplish node prediction, node classification and link prediction making use of features information extracted from graph-structure data. The training process of GCN is shown in Figure 5 below. GCN takes the graph data as inputs, and performs a convolution on each node's neighbors, and updates the node with the result of convolution to realize the aggregation of information among nodes. The result is then convolved as the second layer input and the above operation is repeated until the number of layers reaches the expected depth. Finally, GCN can also convert node states into task-related labels, etc. as outputs.

There are two ways to extract graph features, one is though spectral domain, the other is though vertex domain. The spectral domain uses the eigenvalues and eigenvectors of the Laplacian matrix of the graph to study its properties, the vertex domain finds out the neighbor nodes adjacent to each vertex, defines the connection relationship between the nodes, and then aggregates the information of the neighbor nodes.

(1) Spectral-based Graph Convolutional Neural Networks

The Laplacian matrix of the graph is written as $L = D - A,$ where D refers to the degree matrix of nodes and A refers to the adjacency matrix of nodes. GCN is based on Laplace's spectral decomposition, GCN decomposes a matrix into the product of eigenvalues and eigenvector matrices, In this equation, U is the orthogonal matrix composed by the unit eigenvectors. As shown in Equation (4).

$$L = U\Lambda U^{-1} = U\Lambda U^{T} \tag{4}$$

The first graph convolution model proposed by Bruna is as follows,but the $y_{output=\sigma(Ug_\theta(\Lambda)U^T x)}$original graph convolution neural network has a large amount of computation, but it is of poor performance due to massive calculation and lack of spatial locality. In order to increase the spatial locality of the first-generation graph convolution model, De fferard et al. [14] proposed the second-generation graph convolution neural network in 2016 as in Formula (5).

$$y_{output=\sigma \sum_{j=0}^{K} a_j L^j} \tag{5}$$

It not only improved kernel convolution, but also further simplify the calculation. After several improvements since GCN was proposed, T.N.Kipf and M.Welling et al. [15] put forward the third-generation graph convolutional neural network in 2017,which deepened the depth of the network and defined the propagation mode between layers as shown in Formula (6) .

$$H^{(l+1)} = \sigma\left(\widetilde{D}^{\frac{-1}{2}}\tilde{A}H^{l}\widetilde{D}^{\frac{-1}{2}}W^{l}\right) \tag{6}$$

In the equation, $\tilde{A} = A + I, \widetilde{D}$is the degree matrix of$\tilde{A}$,H Is the characteristic of each layer, for the input layer, W is the parameter matrix, σ is the nonlinear activation function.

(2) Vertex-based Graph Convolutional Neural Networks

Graph convolution based on vertex domain can be likened to the convolution operation in Euclidean space. Starting from vertex domain, each central node and its neighbors are aggregated by defining an aggregation function. It can be seen that the problem of vertex domain graph convolution lies in the different number of neighbor nodes of each node in the graph structure. In order to solve this problem, it is necessary to define a convolution kernel that can handle neighbor nodes of arbitrary length, so that the convolution kernel can perform convolution adaptively according to the number of neighbor nodes of different nodes, so as to further extract the features of nodes in the graph. Sum the hidden states of all neighbor nodes, update the current node hidden states, and realize parameter less convolution: the feature representation of the layer, representing the set of neighbor nodes of a node. Graph convolution based on vertex domain can deal with large scale graph structures and is widely used.

4.3.2 Figure Attention network

The method of GCN to obtain the features of the graph space is very dependent on the graph structure, and the weight of different nodes in the neighborhood is the same. In 2018, Yoshua Bengio et al. proposed a Graph Attention network (GAT) model combining attention mechanism and graph convolutional neural network. Graph attention network has two main advantages. One is that different weights can be assigned to each node, and the second is that after introducing attention mechanism, node information is only related to its neighbor nodes, without obtaining the information of the whole graph. Assume that the graph has n nodes and take the set of node feature vectors as input, each node has F features. After being processed, the dimension of the feature vector may change. For both nodes, a linear transformation is used to transform the dimensional features into dimensional features. For a node, calculating the similarity coefficient as $e_{ij} = a\left(W\vec{h}_i, W\vec{h}_j\right)$ and each neighbor node is a shared attention mechanism, which can map the concatenated vector to the real number. Finally, after it is normalized by Soft-max, the attention coefficient can be obtained as shown in Formula (7).

$$a_{ij} = \frac{exp\left(LeakyReLU\left(e_{ij}\right)\right)}{\sum_{k \in N_j} exp\left(LeakyReLU\left(e_{ij}\right)\right)} \tag{7}$$

In the second step, weighted summation of features is carried out, and the feature vector of fusion domain information of each node is output after nonlinear activation function. As shown in Figure 6, the graph attention network learns the attention weight between two nodes, and performs weighted average on the feature representation of nodes based on the attention weight, so that the feature representation of node 1 can be obtained.

4.3.3 Extensible graph network

The convolution done in GCN incorporates the information from the whole Graph and the efficiency of the GCN can be low if the graph structure is large and the number of nodes is high, which leads to the emergence of the scalable graph network Graph-SAGE [16] with the learning process shown in Figure 7 below. The network is sampled by randomly picking subgraphs and updating the nodes through them, in this way the resulting subgraph structure itself is changing, as a result the model will learn a sampling and aggregation of parameters, avoiding the need to update the node features of the whole graph together during training, thus increasing the scalability.

Figure 7. Graph-SAGE learning process

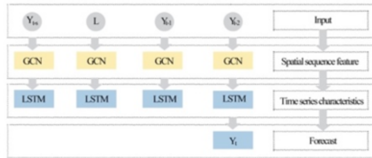


Figure 8. The training flow chart of sales forecasting based on GCN-LSTM

Graph-SAGE first partitions each node to get subgraphs, and then randomly samples some neighbor nodes as feature points for aggregation. After picking out the subgraphs, feature fusion is done, which is able to get the features of the central node in the same way as used by GCN, and finally the nodes are made for classification tasks, etc.

5. GCN-LSTM hybrid model

The sales of goods have time characteristics, and they are related to other spatial characteristics. Graph convolution neural network can be used to process graph data and extract spatial features. LSTM can be used to process time series problems and extract time features of data. Therefore, many scholars have proposed to use the combination of GCN and LSTM to predict traffic flow or air quality, etc. Zhao Ling[17] proposed T-GCN model to predict traffic flow in combination with GCN and GRU, and Qi Bolin [18] used GCN-LSTM model to monitor air quality of small and micro stations. Therefore, we can try to make use of graph convolution neural network model based on the combination of graph convolution neural network and LSTM for sales on the basis of data with both temporal and spatial characteristics.

5.1 Model construction

For commodity sales, if the store is adjacent to the other store, where the sales of the two retailers interact on each other. Therefore, each merchant is considered as a node, connections do exist between merchants if the distance is less than 3km.We try to build a merchant graph, which represents the number of nodes, an edge, and an adjacency matrix, and represents the connectivity between merchants. None connectivity between merchants exist, the distance between the node is zero, the condition is that the distance is 1 from the nearest retailer. Some information related to the sales of goods from the merchant is regarded as the attribute characteristics, forming a feature matrix to represent the number of node characteristics. Node characteristics include the basic attributes of the business circle where the store is located, crowd characteristics, consumption capacity and commodity sales mentioned above. Node attribute characteristics represent time.

Therefore, the GCN-LSTM model learns the following mapping through the graphic structure of the store and the corresponding feature matrix such as predicting future sales. The model is divided into two steps such as GCN and LSTM. The training flow chart is shown in Figure 8 below. The historical data of merchants are as input, GCN aggregation information is used to obtain spatial characteristics, and then input the time series data integrated with spatial characteristics into the LSTM model to capture the time characteristics, we aim at obtaining the final prediction results.

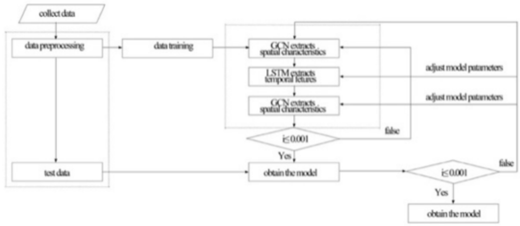


Figure 9. Flow chart about forecasting sales of GCN-LSTM.

The process of GCN-LSTM model about forecasting commodity sales is shown in Figure 9. After preprocessing the input data, the data is divided into training and testing set, If the model does not do well in the test data, we will adjust the parameter of the test model.

5.2 Spatial feature extraction

The key to predict sales is to obtain a spatial structure among businesses. Traditional convolutional neural networks can extract features from spatial data, and deal with the graph structure of locations among businesses. Therefore, graph neural networks are used to process graph data and extract spatial features. GCN can collect the relationship between merchants and their surrounding connected merchants, aggregate the information among merchants, and obtain the spatial correlation of features. In the GCN process, the characteristic matrix containing the merchant information and the adjacency matrix reflecting the location relationship are used as inputs to enter the convolution layer for calculation and aggregation of information. The result of each aggregation will enter the convolution layer again as a new input until the preset number of convolution layers is reached. The final output result is the GCN output that combines the surrounding information and its own information.

5.3 Extraction of time characteristics

Another key to predict the sales of goods is to obtain the time correlation among data. Traditional RNN is limited to long-term prediction and is only suitable for learning short-term memory. As a variant of GNN, LSTM [19] can solve the problems in GNN by using the gating mechanism to remember much long-term information. Therefore, LSTM model is used to obtain time characteristics from sales data, which represents the hidden state of the time, the sales information of the time, the cell state of the time, and the output state of the time.

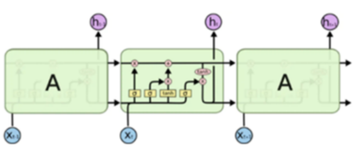


Figure 10. Structure of LSTM [20]

Algorithm 1 The Process of Graph Convolution

Input:
The graph $G(V, E)$;feature matrix X_{t-1};adjacency matrix A;
degree matrix $\hat{D}$;layer number L
$\hat{A} = A + I$
$\hat{D} = \sum_{j=1}^{n} A_{ij}$

```
for i = 1 to n do
    for l = 1 to L do
        H^0_{t-i} = X_{t-i}
        H^{l+1}_{t-i} = σ(D̂^{-1/2} Â D̂^{-1/2} H^l_{t-i} W^l)
    end for
    X̂_{t-i} = H^L_{t-i}
end for
X_t = GRU([X_{t-n}, X_{t-n+1}, ..., X_{t-1}])
```

Output:
X_t

Figure 11. Pseudo code of graph convolution.

LSTM decides which information is discarded from the cell state through the forgetting gate, what is put into the cell state through the input gate, after updating the cell state, output the final value through the output gate. We input the hidden state of the moment and information about the current sales to predict the sales of the next moment. The structure of LSTM is shown in Figure 10.

5.4 GCN-LSTM

In order to predict the sales, GCN-LSTM model is used to simultaneously extract the spatial and temporal characteristics of the data. The specific calculation process is as follows:

$$Ft=\sigma\left(w_f * \left[h_{t-1,}x_t\right] + b_f\right) \tag{8}$$

$$i_t = \sigma(w_i * \left[h_{t-1,}x_t\right] + b_i \tag{9}$$

$$z_t = tanh(w_c * \left[h_{t-1,}x_t\right] + b_c \tag{10}$$

$$c_t=f_t * c_{t-1}+i_t * z_t \tag{11}$$

$$o_t = \sigma(w_o * \left[h_{t-1,}x_t\right] + b_o \tag{12}$$

$$h_t = o_t * tanh\ (c_t) \tag{13}$$

Wherein, the process of graph convolution, the weight and deviation are in the training process. The pseudo code for the graph convolution process in this study is shown in Figure 11 below. The graph convolution process in GCN-LSTM model can make good use of the location structure among adjacent businesses to obtain the spatial characteristics in the data. The LSTM process can capture the characteristics of the dynamic change of sales with time, obtain the time correlation, and finally achieve the prediction of commodity sales.

6. Experimental analysis and conclusion

The work content of this study can automatically extract important features from data to make up for the heavy workload of manual feature search. Potential features among data that cannot be extracted manually, so as to improve work efficiency and realize intelligent and accurate product marketing. However, there are some shortcomings in this research work, such as incomplete data collection and many comprehensive factors which affect product delivery. In the future, we will also build relevant models according to local conditions to create more profound and effective application value in the field of intelligent product marketing.

6.1 Data description and preprocessing

The data of business circle and sales from different brands may be empty when input. In order to facilitate the subsequent learning of neural network, it is necessary to process them.

Table 3. The label of retailer

Basic attribute of business district	**The number of residential areas/The number of shopping centers/The number of office buildings**
Features of the crowd	Resident population/The floating population/Working population/The population of permanent residents/Gender/ age /degree /marriage /status of economy
Consumption level of business scope	In the gear/Low bit/The average gear/Level of consumption/ The dining level
Market status of the business area	Average price of from different brands in recent three months
Indicators about business circle consumption	Sales of different brands in recent March
Consumption preferences in the business circle	Weekly sales of different brands of products

The numerical data should be supplemented with 0, and the categorical data should be supplemented with null. Considering the data set is numeric data type and categories, numeric data under different business circle properties could be orders of magnitude more than the gap, could lead to a neural network to ignore small orders of magnitude, the characteristics of the categories of data due to not numeric, not directly input neural network training, so the need for numeric data type and category data transformation. For numerical data, log1p function operation is performed uniformly, and a relatively smooth and Gaussian distribution data is obtained. Label-Encoder is used to encode the categorical data to obtain the numerical features as shown in table 3.

6.2 Evaluation Index

In order to evaluate the prediction performance of the two algorithms for sales, absolute mean error (MAE) and root mean square error (RMSE) are used to evaluate the deviation among actual sales and predict sales, and accuracy is used to evaluate the effect of the model, as shown in Equation (14-16). The better the prediction is, the closer the prediction is to the real value, and the smaller the deviation is, the smaller the MAE and RMSE will be.

$$\text{MAE}=\frac{1}{n}\sum_{t=1}^{n}|y-\check{y}_t| \tag{14}$$

$$\text{RMSE}=\sqrt[2]{\frac{1}{n}\sum_{t=1}^{n}(y-\check{y}_t)^2} \tag{15}$$

$$\text{Accuracy}=1-\frac{\|y-\check{y}_t\|_F}{\|y\|_F} \tag{16}$$

6.3 The experimental results

In the experiment, shallow neural network and GCN-LSTM models are respectively used to predict the cigarette sales of 10 retail households. The comparison of prediction results is shown in figure 12.

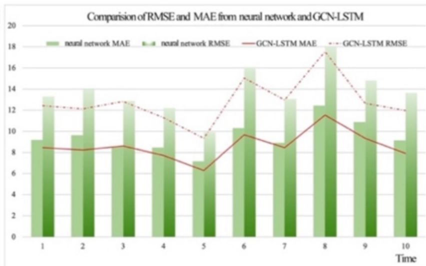


Figure 12. Comparison of RMSE and MAE from neural network and GCN-LSTM

As can be seen from the results in the table, the MAE and RMSE evaluation indexes of the GCN-LSTM model are both smaller than those of the shallow neural network model, indicating that the deviation between the predicted value and the real value of the sales volume of the model is smaller and the prediction effect is better.

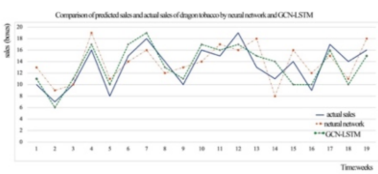


Figure 13. comparison of two methods in 20 weeks

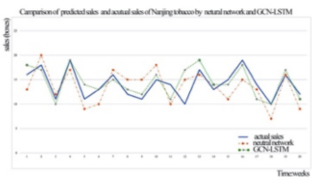


Figure 14. comparison of two methods in 20 weeks

The experiment enumerates the weekly sales forecast and the monthly sales forecast for 10 different retailers respectively, and uses the two algorithms to make the forecast volume experiment in Python.

Figure 13-14 shows the sales sequence diagram of a retail customer's cigarettes from April 2020 to May 2022, as well as the comparison between the actual sales volume and the predicted sales volume of GCN-LSTM from the beginning of 2022, and the prediction accuracy of the two algorithms is compared. As can be seen from the figure, the average prediction accuracy of shallow neural network is 81%, and that of GCN-LSTM is 89%. It shows that the GCN-LSTM algorithm has a better prediction effect on sales volume, and the predicted value is closer to the real value.

6.4 Experimental Conclusion

In conclusion, the neural network algorithm using feature fusion and the GCN-LSTM hybrid model algorithm can make up for the shortcomings of traditional manual cigarette delivery. GCN-LSTM algorithm is closer to the actual sales in the experiment, and an average accuracy is 89%. It can predict the quantity of cigarette marketing accurately, so as to achieve intelligent and automated cigarette marketing and improve the work efficiency of commercial companies. To solve the problem of intelligent precision marketing of goods, this research proposes a shallow neural network algorithm and a GCN-LSTM hybrid model algorithm, which are used to forecast the marketing volume respectively. Under the experimental scenario of forecasting the weekly sales volume, monthly sales in different time periods of Zhen-long and Nanjing, it is verified that the average prediction accuracy of the shallow neural network is 81%, and the average prediction accuracy of GCN-LSTM is 89%. It can be seen that the prediction results of the GCN-LSTM hybrid model algorithm are more consistent with the actual sales data, indicating that the algorithm can improve work efficiency and better control the marketing of goods to different retailers, so as to achieve precise regulation and achieve intelligent marketing. This research can provide reference value for commercial companies in the field of commodity marketing. In the future, more marketing data will be collected and more effective algorithm models will be designed.

7. Conclusion

In order to solve the problem of intelligent precision marketing of goods, this study proposes shallow neural network algorithm and GCN-LSTM hybrid model algorithm. The two intelligent algorithms are respectively used to predict marketing volume. Under the experimental scenario of predicting the sales of Zhenlong (hard Lingyun) and Nanjing (hard Hong) in a week, a month and sales in different time periods, it is verified that the average prediction accuracy of shallow neural network is 81%, and the average prediction accuracy of GCN-LSTM is 89%. It can be seen that the prediction results of GCN-LSTM hybrid model algorithm are more consistent with the actual sales data, which indicates that the algorithm can improve the work efficiency and better control the product marketing to different retail customers, so as to achieve accurate regulation and realize intelligent marketing. This study can provide reference value for commercial companies in the field of commodity marketing. In the future, more marketing data will be collected and more effective algorithm models will be designed.

Acknowledgement

The authors thank ZHANG Mi-Mi, LIU Shen-Shen and GAO Lian-Xin for providing helpful statistical suggestion.

Reference

[1] Li Xian-zong. Research on product recommendation based on double Knowledge Graph and graph convolutional neural network. Master, Jilin University, 2022.
[2] J. Luo Chang-lin, Chen Meng.Research on guidance training strategy of product marketing based on WEFCMO algorithm. Journal of Hebei Academy of Sciences, 2020, 37 (02), 1-7.
[3] Ke-Miao, Huang Hua-guo. E-commerce commodity sales forecasting method based on LSTM neural network. Journal of Fu-qing University, Fujian Normal University, 2020, (05), 83-89.
[4] Han Ya-Juan. Gao Xin. E-commerce product sales prediction based on Machine learning combination model. Application of Computer Systems 2022, 31 (01), 315-321.
[5] Zheng Yan. Qing dynasty. Xiao Y J. Research on forecasting model of commodity demand based on time series. Journal of Chongqing University of Technology (Natural Science) 2019, 33 (09), 217-222.
[6] Huang Wen-Ming. Research on product sales forecasting based on deep learning. Master, Nanjing University of Science and Technology, 2019.
[7] Huang, D., Liu, H., Bi, T., Yang, Q., GCN-LSTM spatiotemporal-network-based method for post-disturbance frequency prediction of power systems. Global Energy Interconnection 2022, 5 (01), 96-107.
[8] Wu Yu. Research on feature fusion mechanism and application of deep neural networks. Ph.D., Sichuan University, 2021.
[9] Tan Ying-Ying, Jun-li wang, Zhang Chao-bo. A survey of text classification methods based on Graph Convolutional Neural network. Journal of Computer Science, 1-14.
[10] Li Jianhui, Zhi-Hong shen, Meng X F. Scientific Big Data management: Concepts, Technologies and systems. Computer Research and Development 2017, 54 (02),
[11] Gori M , Monfardini G, Scarselli F. A new model for learning in graph domains[C]// IEEE International Joint Conference on Neural Networks. IEEE, 2005.

[12] Li Zhang, Heda Song. Nikolaos Aletras, Haiping Lu. Node-Feature Convolution for Graph Convolutional Networks[J] Pattern Recognition,2022.
[13] Bruna J, Zaremba W, Szlam A, et al. Spectral Networks and Locally Connected Networks on Graphs[J]. Computer Science, 2013.
[14] Defferrard M, Bresson X, Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering[J]. Advances in neural information processing systems, 2016, 29: 3844-3852.
[15] Kipf T N, Welling M. Semi-Supervised Classification with Graph Convolutional Networks[C] //2016.
[16] Velikovi P, Cucurull G,Casanova A , et al. Graph Attention Networks[C]// 2017.
[17] Zhao Ling, Song Y, Zhang C, et al. T-GCN: A Temporal Graph Convolutional Network for Traffic Prediction[J]. IEEE Transactions on Intelligent Transportation Systems, 2019, PP (99):1-11.
[18] QI Bailin, Guo Kunpeng,Yang Bin, et al. Air quality prediction based on GCN-LSTM [J]. Application of Computer Systems, 2021.Hochreiter S, Schmidhuber J. Long Short-Term Memory[J]. Neural Computation, 1997, 9(8):1735-1780.
[19] Song Jihong, Wang Tongtong. Encrypted traffic classification method based on feature fusion convolutional neural network. Information and Computer (Theoretical Edition) 2021, 33 (02), 49-51.
[20] Dengyi Huang et al. GCN-LSTM spatiotemporal-network-based method for post-disturbance frequency prediction of power systems [J] Global Energy Interconnection, 2022 (2),96-10.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230166

ICT Framework for the Personal Learning Environment (PLE-ICT) in Higher Education: Results from Experts' Interview

Xiaoshu Xu[a], Xibing Wang[b,1] , Yunfeng Zhang[c], Yan Yue[a] and Yunyang Wu[a]

[a] *School of Foreign Studies, Wenzhou University, China*

[b] *Stamford International University, Thailand*

[c] *Centre for Portuguese Studies, Macau Polytechnic University, China*

Abstract. The incorporation of Information and Communication Technology (ICT) into higher education has profoundly revolutionized the manner in which students engage with educational resources and acquire knowledge. Personal Learning Environments (PLEs) facilitate a tailored and adaptable learning approach, empowering students to take charge of their own educational experiences. Nonetheless, to effectively leverage PLEs in higher education, a robust ICT framework is imperative. This paper explores the development of an ICT framework specifically designed for PLEs within the context of higher education. The framework encompasses diverse ICT attributes, including accessibility, flexibility, data analysis, extensive capacity, automation, personalization, and security. Analyzing interview data led to the formulation of a PLE-ICT framework prototype, comprising five dimensions: ICT hardware, ICT software, ICT services, project development, and support team. The findings of this study enrich the theoretical comprehension of ICT integration in PLEs and offer valuable insights for the future elaboration of more intricate PLE-ICT scales. The suggested ICT framework delivers a comprehensive and pragmatic blueprint for the implementation of PLEs in higher education institutions, with the potential to augment student learning experiences and outcomes.

Keywords. ICT framework, Personal Learning Environment, PLE, higher education, experts' interview

1. Introduction

With the rapid development of information and communication technology (ICT), higher education has also undergone significant changes, leading to the rise of the

[1]Corresponding Author: Xibing Wang (Ph.D. candidate), candidate in Educational Administration and Leadership at Stamford International University, Thailand. E-mail: 2105100003@students.stamford.edu. This work was supported by the China National Social Science Late-stage Funding Project (21FJKB013). The authors declare that there is no conflict of interest.

concept of a Personal Learning Environment (PLE). A PLE is a set of technological tools and resources that an individual learner selects and uses to support their learning activities and goals, in a self-directed and adaptable manner, based on their interests and needs, and with the purpose of building a personal learning network[1]. To construct and implement PLE in higher education, Information and Communication Technology (ICT) plays a crucial role[2]. ICT provides the infrastructure and tools necessary to support learners in managing and organizing their learning experiences and educational resources, including online platforms and digital repositories, social media and communication tools, and personalized learning technologies [3].

Several ICT frameworks and criteria have been proposed to guide the development and implementation of Personal Learning Environments (PLEs) in higher education. One of the most widely recognized frameworks is the Seven Pillars of PLE proposed by Downes [3]. This framework outlines seven critical components of PLEs: ownership and control, aggregation, mixability, reusability, feeding forward, feedback, and potential [3]. Besides, the e-Portfolio Model, as proposed by Attwell [4], emphasizes the role of learners in organizing, reflecting upon, and showcasing their learning experiences and achievements. In this model, ICT tools facilitate the creation and management of digital portfolios, enabling learners to document, track, and share their learning journey with peers, instructors, and potential employers [4]. Moreover, the Connectivist PLE Model, rooted in the connectivism learning theory introduced by Siemens [5], emphasizes the role of networks and connections in the learning process. In this model, learners use ICT tools to access, filter, and evaluate information, as well as to create and share knowledge within their personal learning networks

While the aforementioned PLE models have contributed significantly to the understanding and development of Personal Learning Environments in higher education, there remains a research gap in providing a comprehensive PLE ICT framework that addresses all dimensions necessary for successful construction and implementation. Previous models primarily focused on individual aspects of PLEs, such as personalization, adaptability, self-regulation, collaboration, and networking. However, a holistic PLE ICT framework is needed to ensure seamless integration of all dimensions, including hardware, software, ICT services, project development, and support teams. This research aims to fill this gap by developing a PLE ICT framework that encompasses the complete dimensions of ICT required for constructing and implementing PLEs in higher education. The research question is thus raised:

Question: What are the ICT requirements for the construction and implementation of a PLE in higher education from ICT experts' points of view?

The research context of this study focuses on higher education institutions, where the integration of Information and Communication Technology (ICT) plays a pivotal role in shaping the learning experience. By examining the implementation of Personal Learning Environments (PLEs) within this context, the research seeks to understand how a comprehensive ICT framework can facilitate personalized learning, ultimately leading to enhanced educational outcomes for students.

This study first introduced the ICT resources needed for the construction and implementation of a PLE in higher education, then discussed the concepts and necessity of hybrid cloud, and finally interpreted the ICT characteristics of PLE in higher education. The methodology is then introduced followed by the interview data analysis. The findings of this study will provide insights into the ICT requirements for implementing a PLE and help the future development of a more detailed PLE-ICT scale for higher education. Additionally, the results of this study will also contribute to the growing body of literature on the use of ICT in higher education and the implementation of PLEs.

2. Review of the ICT requirements for the development of PLE in higher education

A comprehensive understanding of the ICT requirements for constructing and implementing Personal Learning Environments (PLEs) in higher education is crucial for their successful integration into the learning experience. Several studies have explored the essential elements and tools that facilitate effective PLEs [6][1]. Accessibility, flexibility, data analysis, large capacity, automation, personalization, and security have been identified as key ICT features for PLEs [7][8].Additionally, recent research emphasizes the importance of designing PLEs that are adaptable to learners' needs, preferences, and learning styles [9][2]. Furthermore, the integration of cutting-edge technologies such as artificial intelligence, machine learning, and cloud computing has been proposed to enhance the effectiveness and efficiency of PLEs in higher education [10]. It is also crucial to consider the role of social media, collaboration tools, and multimedia resources in supporting the development of connected and engaging learning experiences [11]. In sum, the literature highlights the importance of a well-rounded ICT framework that incorporates various technological dimensions and features to ensure the effective construction and implementation of PLEs in higher education.

2.1 ICT resources needed for PLE in higher education

In order to implement PLEs effectively, ICT resources such as laptops, tablets, smartphones, and other mobile devices are required to provide students with access to the digital learning materials and tools they need to complete their studies. Furthermore, software such as learning management systems (LMS), e-portfolios, social software, and content management systems are necessary to manage, store and share learning materials and to facilitate collaboration among students and teachers.

In addition, communication technologies such as instant messaging, blogs, wikis, and discussion forums are crucial for fostering collaboration and communication among students and teachers. These technologies allow students to exchange ideas and provide feedback on their learning experiences in real-time. By analyzing the most popular online learning platforms, the following part introduced the hardware, software, and service needed for implementing PLE in higher education:

2.1.1 Hardware resources

The hardware source needed for personalized learning environments generally includes both mobile and PC. To ensure smooth system operation, the mobile side can use at least an eight-core processor with memory references ranging from 4 GB to 64 GB

RAM. The operating system only needs to support the latest version of iOS or Android, and the network only needs to have a high-speed internet connection to ensure fast loading and syncing of platform data and resources. The PC side can use at least an Intel Core i5 or similar processor with memory ranging from 8 GB to 256 GB RAM, an SSD notebook or desktop computer, and a display screen with at least 1920x1080 resolution for a good visual experience. The network only needs to have a high-speed internet connection to ensure fast loading and syncing of platform data and resources.

2.1.2 Software resources

An effective ICT framework for PLE in higher education should include software resources that support the learning and development needs of students and educators. Key features to consider in educational software include:

- Well-rated apps for both iOS and Android platforms, allowing students to access the software from their preferred device, including laptops, Chromebooks, and smartphones.
- Automated grading and integration with SpeedGrader to support the assessment process.
- A website builder and customizable templates to facilitate the creation of educational content by educators.
- Project creation tools, including a video editor, plagiarism checker, and assignment tracker to support student learning and engagement.
- Integration with corporate social media accounts.
- Tools for authoring, including text editing and video capture support the creation of multimedia learning materials.

2.1.3 Service resources:

In addition to software resources, an ICT framework for PLE in higher education should also include reliable and user-friendly service resources to support the learning process. Key features to consider include:

- Support for third-party cloud storage providers, such as Dropbox, to ensure the safety and accessibility of educational data.
- The ability to create individualized student pathways via LinkedIn Learning content.
- A white-label approach to foster customization and consistent branding across the educational institution.
- Support for multiple tenants, allowing an administrator to manage multiple accounts with a single set of sign-on credentials.

- Advanced support options, including a forum or FAQ, a robust community forum, and growing scalability.

- An easy-to-use interface with a variety of support options, including direct support via email and phone.

2.2 Hybrid cloud

The hybrid cloud concept has gained significant traction across various industries, including education, as it combines the advantages of both public and private clouds. Hybrid cloud computing, as defined by Srinivasan[12], comprises a mix of on-premises, private cloud, and third-party public cloud services with orchestration and integration between platforms. Key components of modern hybrid cloud architecture include virtualization, data-driven approaches, Software as a Service (SaaS), Industry 4.0 integration, on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured service [13][14][15][16][10].

Hybrid cloud offers benefits such as cost efficiency, scalability, flexibility, and security. In the context of PLEs, a hybrid cloud allows institutions to balance cost-effectiveness and scalability with privacy and security for sensitive student data. It also offers flexibility and adaptability, enabling institutions to choose appropriate cloud services and switch between them as needs evolve. Moreover, it enables the integration of advanced technologies such as artificial intelligence and machine learning, enhancing student learning experiences.

The use of mobile devices in higher education is also becoming increasingly popular, with studies showing the potential for mobile devices to improve access to educational resources and facilitate learning [17]. The integration of mobile devices with hybrid cloud infrastructure can further enhance the adaptability and accessibility of PLEs, providing learners with greater opportunities for personalized learning experiences.

The unified hybrid cloud, still in its early adoption phase, demonstrates benefits including scalability, cost efficiency, improved security, increased agility, better compliance, and enhanced collaboration. Papadakis et al. [18] highlight the potential of digital eLearning educational tools, such as ARION, to synchronize, compose, and orchestrate learning session data within a hybrid cloud environment. Although the benefits of a unified hybrid cloud platform depend on an organization's specific needs and the level of integration between different cloud platforms, it holds great potential for optimizing resources and enhancing collaboration across various industries, including higher education.

2.3 The ICT features of PLE in higher education

A review of the literature indicated the following five ICT features of PLE:

- Customizable User Interfaces (CUIs): Manca and Paternò [19] defined CUIs as user interfaces that can be modified according to the needs and preferences of different users. Specifically, CUIs allow users to customize not only the layout and content of the user interface but also its behavior and functionality. In PLE, CUIs can include selecting and arranging the tools and resources that they need to support their learning, such as course materials, discussion forums, and multimedia resources.

- Social Networking Tools: Social networking tools, such as blogs, wikis, and discussion forums, play an important role in PLE, allowing learners to engage with each other, share information, and collaborate on projects [20].

- Personalized Learning Paths: PLEs can provide learners with the ability to create and manage their own learning paths, enabling them to select and prioritize the resources and activities that best meet their learning needs [21]. This can help to increase learner engagement and motivation, as well as support the development of personalized learning plans that are tailored to each individual's needs and interests.

- Mobile Access: With the increasing prevalence of mobile devices, PLEs need to be designed to support access from a variety of mobile devices, including smartphones and tablets [22]. This can provide learners with the ability to access learning resources and engage in learning activities at any time, from any location, and on any device.

- Learning Analytics: Learning analytics tools play an important role in PLE, providing learners with the ability to track and monitor their learning progress, and receive personalized feedback and support [11]. These tools can also be used to analyze learner data and provide insights into their learning behavior and patterns, helping educators to identify areas for improvement and make more informed decisions about their teaching practice.

3. Method

This study employed semi-structured interviews to explore the essential ICT components and systems for constructing effective and efficient PLE-ICT in higher education. This data-gathering method offers a flexible approach, enabling researchers to adapt questions based on participants' responses, generating in-depth and context-specific insights [23]. Semi-structured interviews are particularly useful when investigating complex phenomena like ICT components in PLEs, as they allow researchers to better understand participants' experiences, perceptions, beliefs, and attitudes [24]. This method is well-suited for exploring ICT requirements for PLEs in higher education, as it helps researchers gather nuanced information and identify commonalities and differences among participants' experiences and opinions [25].

The sampling technique utilized was purposive sampling, selecting five ICT experts with 5-10 years of working experience in PLEs in higher education. A sample size of five interviewees was chosen to ensure a manageable amount of data for in-depth analysis while still providing rich and diverse perspectives on the topic. The diversity of participants in gender, age, and professional background aimed to provide a comprehensive understanding of the ICT components and systems necessary for building an effective and efficient PLE, and to ensure that the resulting framework would be applicable to a wide range of higher education institutions.

The purpose of the interviews was explained to the participants, and they all agreed to have the interview recorded. The average interview time was 30 minutes. The interviews aimed to gather data about the ICT components and systems necessary for building an effective and efficient PLE. The collected data was analyzed and coded to identify common themes and patterns.

Based on the literature review, the interview questions were designed to elicit information about the participants' experiences with PLEs, their understanding of the ICT infrastructure required for PLEs, and their recommendations for an ICT framework that would support PLEs in higher education. The questions were open-ended, encouraging participants to provide detailed information and insights about their experiences with PLEs and their thoughts on the ICT infrastructure necessary to support PLEs (see table 1)

Table 1. Question items and the literature reference

Question items	**literature**
1. Please describe your experience with PLEs in higher education.	[26]
2. In your opinion, what are the key components of an ICT framework for PLEs in higher education?	[27]
3. What factors should be considered when implementing an ICT framework for PLEs in higher education?	[28]
4. What challenges or limitations have you encountered, or do you anticipate, in implementing an ICT framework for PLEs?	[29]
5. How do you ensure that ICT features in PLEs are accessible to all learners, including those with disabilities and different backgrounds?	[30]
6. How does flexibility in ICT features contribute to the success of PLEs in higher education?	[9]
7. How does data analysis support the development of effective learning experiences in PLEs?	[8]
8. Why is having a large capacity important for the success of PLEs in higher education?	[7]
9. How does automation support the development of more flexible and dynamic learning environments in PLEs?	[31]
10. How do you ensure the personalization of functions within PLEs in higher education?	[6]
11. Please share any best practices you have observed or implemented for the successful integration of an ICT framework in PLEs in higher education.	Self-made
12. What recommendations do you have for institutions seeking to implement an ICT framework for PLEs?	[32]

The data collected from the interviews were transcribed and analyzed using a thematic analysis approach. The transcripts were read multiple times to gain a general understanding of the data. Codes were then generated and applied to the data to identify common themes and patterns. The codes were developed based on the interview questions and the collected data. The data analysis aimed to identify the ICT components and systems necessary for building an effective and efficient PLE.

4. Sample and sampling

This study aimed to develop an ICT framework for Personalized Learning Environments (PLEs) in higher education. To achieve this goal, the study employed a qualitative research design and conducted in-depth interviews with 5 ICT experts with 5-10 years of working experience in PLEs in higher education. The participants were selected based on their professional experience and expertise in the field of ICT and education. All participants held a Ph.D. degree, with two of them being female and three of them being male. Their ages ranged from 35-55.

Before conducting the interviews, informed consent was obtained from all interviewees. They were provided with a written description of the study, including its purpose, methods, and expected outcomes. Participants were also informed of their right to withdraw from the study at any time. They were required to sign a written consent form indicating their agreement to participate in the study. The transcripts of the interviews were coded using a pseudonym to protect the identity of the participants.

5. Results and discussion

The transcribed data from the interviews were analyzed using a thematic coding approach, following a multi-stage process. Initially, the data were subjected to open coding, where common themes and patterns were identified [33]. Next, these themes were grouped into similar ideas and labeled, resulting in the creation of categories [34]. To ensure consistency and accuracy, two coders conducted the coding process independently, and their results were compared. Any discrepancies were discussed and resolved through consensus [35].

Coding categories: The following coding categories were identified and used for analysis.

- Key components of an ICT framework for PLEs
- Considerations for the implementation of an ICT framework for PLEs
- Challenges and limitations in the implementation of an ICT framework for PLEs
- Best practices for the successful implementation of an ICT framework for PLEs

The coded data were analyzed to identify common themes and patterns in the data. The findings from the analysis were used to develop a comprehensive ICT framework for PLEs in higher education, which involved the following seven ICT features of PLE:

1) Accessibility

It can refer to physical accessibility, such as access to the necessary hardware and software, as well as to the usability and user-friendliness of the ICT features themselves [30]. It is also important to consider accessibility in terms of socio-economic factors, such as the cost of access and the availability of technical support [36]. To ensure that ICT features in PLEs are accessible to all learners, it is necessary to consider the needs of different groups and to design the features to be inclusive and user-friendly [30].

2) Flexibility

It refers to the ease and convenience of ICT deployment, as well as its ability to be recovered and monitored. Moreover, it also involves the ability to adapt to the needs and preferences of each individual learner [9]. This requires the provision of reliable and efficient infrastructure, as well as the development of user-friendly systems and tools [36].Besides effortless implementation, it is essential that ICT components within PLEs can be readily restored in the event of system failures or other technical problems. Finally, ICT features in PLEs should be designed to be easily monitored, allowing for the effective measurement of their impact on learning and for the identification of areas for improvement [36].

3) Data analysis

It enables PLEs to provide detailed information about learners' behavior and learning strategies, as well as to support the continuous improvement of learning experiences [8]. Data analysis in PLEs can include the analysis of learning behavior, such as the types of resources used and the strategies employed to engage with the material [37]. This information can be used to support the development of more effective learning experiences, as well as to identify areas for improvement and track progress over time [8].Furthermore, analyzing data in PLEs can offer significant insights into elements that impact learning, such as motivation, engagement, and self-confidence [37]. This information can be used to support the development of more effective teaching and learning strategies, as well as to identify areas where additional support may be needed to ensure that all learners are successful [8].

4) Large capacity

The ability of an online learning system to handle a large volume of users and data. This could include the ability to process user data and generate personalized recommendations, as well as the ability to store and manage a large number of learning resources [38]. By tracking and storing learners' historical behavior data, personalized recommendation systems can use machine learning algorithms to analyze this data and provide personalized recommendations to learners, helping them achieve their learning goals more effectively [7]. The capacity to save and retrieve learning documents, including notes, assignments, and evaluations, plays a crucial role in creating a significant learning portfolio. The large storage capacity of ICT in PLEs also allows learners to easily access and manage learning resources, such as readings, videos, and multimedia materials [38].

5) Automation

It refers to the use of machine learning algorithms to personalize the learning experience of students without the need for manual intervention by instructors [31].

This feature helps to streamline these processes, making them more efficient and reducing the administrative burden on educators and learners. The use of automated systems in PLEs can also support the development of more flexible and dynamic learning environments.

6) Personalization of functions

It refers to the use of various technologies and methods to adapt learning experiences and materials to individual learners, taking into account their learning styles, interests, and prior knowledge [6]. This feature can support the development of more personalized and effective learning experiences, as learners can tailor their learning environment to their specific needs. The personalization of ICT functions can also support the development of more flexible learning environments, as learners can access and use the tools and resources that best meet their needs and preferences [39].

Security

It includes protecting sensitive student and institutional data. According to a study by Garg and Goel [40], the threats to online assessment security include identity fraud, cheating, and hacking. Security strategies such as multi-factor authentication, proctoring, and machine learning-based approaches are recommended. Furthermore, security measures can prevent unauthorized access to ICT systems, reducing the risk of data breaches and theft of confidential information. It should be pointed out that ICT systems in higher education must comply with various regulations which set strict guidelines for the handling of student records and information.

Based on the literature review and interview data analysis, the researchers constructed the following PLE-ICT framework as a prototype (see Table 2).

Table2. PLE-ICT framework prototype

1. Hardware	**1**	**2**	**3**	**4**	**5**
1.1 Have a Standard Operating Procedure (SOP)					
1.2 Inspection and improvement, adjustment of the SOP					
1.3 User feedback mechanism					
1.4 Expert feedback mechanism					
2. Software					
2.1 Have a Standard Operating Procedure (SOP)					
2.2 Inspection and improvement, adjustment of the SOP					
2.3 User feedback mechanism					
2.4 Expert feedback mechanism					
3. ICT service					
3.1 Have a Standard Operating Procedure (SOP)					
3.2 Inspection and improvement, adjustment of the SOP					

3.3 User feedback mechanism					
3.4 Expert feedback mechanism					
4. Project development					
4.1 Have a Standard Operating Procedure (SOP)					
4.2 Inspection and improvement, adjustment of the SOP					
4.3 User feedback mechanism					
4.4 Expert feedback mechanism					
5. Supporting group					
5.1 Have a Standard Operating Procedure (SOP)					
5.2 Inspection and improvement, adjustment of the SOP					
5.3 User feedback mechanism					
5.4 Expert feedback mechanism					

1	Not implemented: no standardized operating procedures or user and expert feedback mechanism.
2	Planning: plans to establish standardized operating procedures and a user and expert feedback mechanism.
3	Partially implemented: partially in accordance with the standardized operating procedures, with a basic user and expert feedback mechanism
4	mostly implemented: mostly in accordance with the standardized operating procedures, with a relatively complete user and expert feedback mechanism.
5	Fully implemented: fully in accordance with the standardized operating procedures, with a complete user and expert feedback mechanism.

Source: self-generated

As shown in the above table, the main technical indicators of the Personal Learning Environment include five dimensions: ICT hardware, ICT software, ICT services, project development and support team.

Hardware equipment is the foundation for platform operation and is crucial for the efficient and stable operation of the platform [14]. ICT hardware includes servers, storage and bandwidth, and networks. The server is the core of the entire system and is responsible for processing user requests, providing content to users, and personalized services [10]. Storage devices are used to store all user data and content and need to have a large capacity and high read and write speed to ensure that users can access the content they need at any time and anywhere [39]. Bandwidth and network are the communication channels between the platform and users and determine the network speed and stability of users when using the platform.

Software systems are an essential component in personalized learning environments [41]. They are responsible for implementing various functions in the learning environment, ensuring that both students and teachers can learn in an efficient and stable environment [7]. The ICT software includes systems, tools, and backend management, which together form an efficient and stable learning environment.

Information Technology Service Management (ITSM) is a key factor in personalized learning environments [40]. Its use helps to ensure system reliability, improve service efficiency, and enhance user experience. The main service of ICT refers to ITSM, such as ISO15000, which provides a systematic process and method for service management to ensure service reliability and availability.

The development team is an important support force for the personalized learning platform [16]. It provides users with high-quality technical and service support, ensuring that the platform is always in its best state to better support the users' learning needs. The ICT support team primarily provides technical support such as equipment, video filming, and instructional material development to teachers, students, and administrative staff.

In sum, creating and utilizing PLE in higher education requires attention to five main areas: ICT hardware, software, services, project development, and support team. Hardware such as servers, storage devices, and networks are essential for the PLE's efficient operation. The software implements various functions, ensuring a smooth experience for students and teachers. IT Service Management (ITSM) improves reliability and efficiency, while the project development team ensures the platform meets users' needs. Technical support provides assistance to teachers, students, and staff, ensuring optimal performance. Overall, these dimensions are crucial for constructing and implementing an effective PLE in higher education.

6. Conclusion and limitation

This study presents the development of an Information and Communication Technology (ICT) prototype framework for Personal Learning Environments (PLEs) in higher education from the perspective of ICT experts. The framework considers various ICT features, including accessibility, flexibility, data analysis, large capacity, automation, personalization, and security, as well as the five dimensions of ICT hardware, ICT software, ICT services, project development, and support team. The construction of this ICT framework for PLEs is a significant advancement in providing personalized and flexible learning opportunities in higher education. The proposed PLE-ICT framework provides a prototype framework from the perspective of ICT experts, which offers insights for future detailed PLE-ICT scale development and supports institutions looking to enhance personalized learning experiences in higher education.

Based on the findings, practitioners are recommended to consider the identified ICT features and dimensions when implementing PLEs in higher education settings. This will help ensure that PLEs are accessible, flexible, and capable of providing personalized learning experiences for students.

The study has some limitations: first, only five ICT experts were interviewed, which may not accurately represent the views and perspectives of other stakeholders in higher education, such as teachers and students; second, the semi-structured interview relied on participants' self-reported information, which may be subject to biases and limitations in their perceptions and recall of events. Future research could collect empirical data to enrich and test the PLE-ICT framework.

References

[1]Castañeda L, Tur G, Torres-Kompen R, Attwell G. The influence of the personal learning environment concept in the educational research field: A 2010-2020 systematized review. International Journal of Virtual and Personal Learning Environments (IJVPLE). 2022;12(1):1-16.

[2]Shemshack A, Spector J M. A systematic literature review of personalized learning terms. Smart Learning Environments. 2020;7(1):1-20.

[3]Downes S, E-Learning 2.0. International Journal of Instructional Technology and Distance Learning. 2005;2(1):1-6.

[4]Attwell G. Personal Learning Environments-the future of eLearning. Elearning papers. 2007;2(1):1-8.

[5]Siemens G. Connectivism: A learning theory for the digital age. International Journal of Instructional Technology and Distance Learning. 2005;2(1):3-10

[6]Alamri H A, Watson S, Watson W. (2021). Learning technology models that support personalization within blended learning environments in higher education. TechTrends. 65:62-78.

[7]Fan J, Jiang Y, Liu Y, Zhou Y. Interpretable MOOC recommendation: a multi-attention network for personalized learning behavior analysis. Internet Research. 2022;32(2):588-605.

[8]FitzGerald E, Jones A, Kucirkova N, Scanlon E. A literature synthesis of personalised technology-enhanced learning: what works and why. Research in Learning Technology. 2018;26.

[9]Mesquida A D, Ferrer G T, Garcias A P, Moral S V. Flexible Learning Itineraries in Digital Environments for Personalised Learning in Teacher Training. RIED. Revista Iberoamericana de Educación a Distancia. 2022;25(2):173-188.

[10]Xin J, Zazueta F. Technology trends in ICT–towards data-driven, farmer-centered and knowledge-based hybrid cloud architectures for smart farming. Agricultural Engineering International: CIGR Journal. 2016;18(4):275-279.

[11]Pachler N, Bachmair B, Cook J. Mobile learning: Structures, agency, practices. Springer Science & Business Media. 2009.

[12]Srinivasan, A., Quadir, M. A., & Vijayakumar, V. (2015). Era of cloud computing: A new insight to hybrid cloud. Procedia Computer Science, 50, 42-51.

[13]Mell P, Grance T. The NIST Definition of Cloud Computing. National Institute of Standards and Technology (NIST), US Department of Commerce. 2011.

[14]Moreno-Vozmediano R, Montero R S, Llorente I M. Iaas cloud architecture: From virtualized datacenters to federated cloud infrastructures. Computer. 2012;45(12):65-72.

[15]Satyanarayana S. Cloud computing: SAAS. Computer Sciences and Telecommunications. 2012;(4):76-79.

[16]Sundarakani B, Kamran R, Maheshwari P, Jain V. Designing a hybrid cloud for a supply chain network of Industry 4.0: a theoretical framework. Benchmarking: An International Journal. 2021:28(5);1524-1542.

[17] Sung, Y. T., Chang, K. E., & Liu, T. C. (2016). The effects of integrating mobile devices with teaching and learning on students' learning performance: A meta-analysis and research synthesis. Computers & Education, 94, 252-275.

[18]Papadakis, A., Barianos, A., Kalogiannakis, M., Papadakis, S., & Vidakis, N. (2022). ARION: A Digital eLearning Educational Tool Library for Synchronization Composition & Orchestration of Learning Session Data. Applied Sciences, 12(17), 8722.

[19]Manca M, Paternò F. Customizable dynamic user interface distribution. In Proceedings of the 8th ACM SIGCHI Symposium on Engineering Interactive Computing Systems. 2016 June;27-37.

[20]Downes S, Models for sustainable open educational resources. International Review of Research in Open and Distance Learning. 2007;8(1):1-13.

[21]Castells, M. (2009). Communication Power. Oxford, Oxford University Press

[22]Johnson L, Adams Becker S, Estrada V, Freeman A. The NMC/CoSN horizon report: 2012 K-12 edition. Austin, TX: The New Media Consortium. 2012.

[23]Jamshed S. Qualitative research method-interviewing and observation. Journal of Basic and Clinical Pharmacy. 2014;5(4):87-88.

[24]Rubin H J, Rubin I S. Qualitative interviewing: The art of hearing data. sage. 2011.

[25]Galletta A. Mastering the semi-structured interview and beyond: From research design to analysis and publication (Vol. 18). NYU Press. 2013.

[26]Dabbagh N, Kitsantas A. Personal Learning Environments, social media, and self-regulated learning: A natural formula for connecting formal and informal learning. The Internet and higher education. 2012;15(1):3-8.

[27]Fiedler S H, Väljataga T. Personal learning environments: concept or technology?. International Journal of Virtual and Personal Learning Environments (IJVPLE). 2011;2(4):1-11.

[28]Chatti M A, Jarke M, Frosch-Wilke D. The future of e-learning: a shift to knowledge networking and social software. International journal of knowledge and learning. 2007;3(4-5):404-420.
[29]Redecker C. Review of learning 2.0 practices: Study on the impact of Web 2.0 innovations on education and training in Europe. 2009.
[30]Khetarpal A. Information and communication technology (ICT) and disability. Review of market integration. 2014;6(1):96-113.
[31]Prihar E, Haim A, Sales A. Heffernan N. Automatic Interpretable Personalized Learning. In Proceedings of the Ninth ACM Conference on Learning@ Scale. 2022, June;1-11.
[32]Kop R, Hill A. Connectivism: Learning theory of the future or vestige of the past?. International Review of Research in Open and Distributed Learning. 2008;9(3):1-13.
[33]Braun V, Clarke V, Using thematic analysis in psychology. Qualitative Research in Psychology. 2006;3(2):77-101. https://doi.org/10.1191/1478088706qp063oa.
[34]Fereday J, Muir-Cochrane E. Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. International journal of qualitative methods 2006;5(1):80-92.
[35]Creswell J W, Poth C N. Qualitative inquiry and research design: Choosing among five approaches. Sage publications. 2016.
[36]Warschauer M, Matuchniak T. New technology and digital worlds: Analyzing evidence of equity in access, use, and outcomes. Review of Research in Education. 2010;34(1):179-225.
[37]Gee J P. What video games have to teach us about learning and literacy. Computers in Entertainment (CIE). 2003; 1(1):20-20.
[38]Wei X, Sun S, Wu D, Zhou L. Personalized online learning resource recommendation based on artificial intelligence and educational psychology. Frontiers in psychology. 2021;12:767837.
[39]Kurilovas E, Kubilinskiene S, Dagiene V. Web 3.0–Based personalization of learning objects in virtual learning environments. Computers in Human Behavior. 2014;30:654-662.
[40]Garg M, Goel A. A systematic literature review on online assessment security: Current challenges and integrity strategies. Computers & Security. 2022;113:102544.
[41]Rani M, Nayak R, Vyas O P. An ontology-based adaptive personalized e-learning system, assisted by software agents on cloud storage. Knowledge-based systems. 2015;90:33-48.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230167

An Empirical Approach to the Utilization of Affective Decision Tree Models in Smart Teaching

Ran Li[a,1]
[a] *School of English Language and Culture, Xi'an FanYi University, Xi'an, China*

Abstract. In the upcoming age of AI, smart education reshapes not only the macro modality of society and education but also the micro modality of knowledge and learning. The top priority of smart education is the timely learning disability diagnosis and feedback, the assessment and improvement of students' mental quality, in which affective data is the crucial index. Putting in perspective the working of positive and negative affects in smart teaching, this research proposes the affective decision tree model drawing on the affective data collected from the PANAS test. Based on the C5.0 algorithm, classification rules are extracted from the calculation of the information gain ratio of affective variables and the construction of decision tree is realized through SPSS 26.0. The research findings indicate that 1) negative affects took precedence over positive affects in the formulation of participants' smart learning strategy; 2) the architecture of the affective decision tree constructed via learners' affective judgments and ratings represents the affective filtering process which optimizes teachers' selection and organization of smart teaching and learning activities. Accordingly, the affective decision tree model could be used as the efficient prediction model and affective profiling model and teaching support system in smart teaching, prompting the affective and cognitive connectivity between teachers and students.

Keywords. smart teaching, affective decision tree, educational data mining, PANAS test, CART algorithm

1. Introduction

The booming of smart technologies like Cloud Computing, Big Data, IoT, Artificial Intelligence and Mobile Internet delineate a new vision for smart education[1]. The essence of smart education lies in the construction of an technology-energized ecological learning environment which facilitate learner's ubiquitous and self-adaptive learning aided by the smart teaching paradigm driven by educational data and man-machine coordination, focusing on the precision in evaluation and management[2].

[1] Corresponding author, Ran Li, School of English Language and Culture, Xi'an FanYi University, Xi'an, China; Email: Email: 542501525@qq.com. Research Interest: cognitive psychology, educational technology.
This work was supported by Xi'an Social Science Foundation project entitled "Research on the Current Situation and Intervention Mechanism of Online Learning Burnout of College Students in Xi'an under the Background of Smart Education" (Project Number: 22JY51).

Geared to learners' developmental status, smart teaching features situational awareness technologies such as natural language understanding, expression recognition, affective computing and learning analytics technologies such as educational data mining, student profiling to achieve the intelligent diagnosis and management of students learning state[3].

Currently, there are two mainstream interdisciplinary approaches to smart teaching, one is the more fruitful behavioral approach benefiting from the data-mining of the learning process and performance, the other is the more prospective profile approach recognizing and visualizing the delicate interaction between the macro cognitive systems, motives and affective systems in learners' personality[4]. The second approach is booming as the result of the advancement in affective computing which features the development of hardware and software framework for the measurement of student's affective and cognitive states based on some sensing system. One of the cutting-edge research orientation in the interdisciplinary field between smart teaching and affective computing is the construction of an adaptive emotionally-intelligent e-learning system which is the enabler of not only affect recognition and analysis but also the simulation the traditional classroom's pedagogy in the system on the basis of learning psychology regulation. In the affect-oriented smart classroom, learners' multi-modal data extracted from the use of expressions, speech, eye movement, and physiological signals, as well as text and questionnaires has been used to infer students' emotional state and learning state and affective computing has been applied to stress detection, dropout prevention and teaching resources optimization.

However, previous data mining models applied in affective computing invariably focused on the process data or environmental data, too much weight has been laid on the smart learning environment of augmented reality or the reconfiguration of all the structural elements defining a smart classroom, the data revealing learners' cognitive and emotional need and state have been inadequately exploited. Therefore, to bridge smart teaching and smart learning via data mining and learning analytics, the present study aims to construct a decision tree model based on affective attributes and information entropy, revealing the intrinsic correlation between the working of affect mechanism, learners' self-efficacy and motivation and teachers' smart teaching decision through an empirical PANAS (Positive and Negative Affect Schedule) test to recommend the optimal smart teaching and learning strategies.

2. The decision tree models in educational data mining

EDM (Educational Data Mining) concerns the extraction and exploration of useful information from educational data. In EDM, decision trees are valuable predictive and descriptive analytic tool that visualize the correlation between input variables. The major functions of decision trees in EDM are listed as follows: 1) classify observations based upon a target that is binary, ordinal, or nominal; 2) predict an outcome for an interval target; 3) predict a decision when you identify the specific decision alternatives.

Decision trees can be regarded as the distribution of conditional probability in feature space and class space. Normally, the classification rules split classification trees into nodes and directed edges with. Nodes generally belong to two major types, internal nodes, designating a feature or attribute, and leaf nodes, designating a class or category.

The value of variables on the internal nodes determines the branching scope of a decision tree. Starting from the root node, a series of statistical significance tests will be conducted to allocate a specific feature of the instance into certain nodes based on a reduction in entropy, variance, or Gini impurity, each internal node corresponding to a specific value of the feature. Through the recursive tests and allocation, the instances eventually reaches the relevant classes of the leaf nodes[5].

Recently, relevant researches in the field of machine learning suggest decision trees can be used as effective predictive models in the predictive analysis of placement of students[6], students' study path selection[7], detecting undesirable students' behaviors and evaluating students' performance in an E-learning environment[8][9]. Likewise, as an effective data-mining tool, decision trees can also be used in online learning behavior feature mining[10] and the computation of students' academic performance[11]. Noticeably, scholars have been using decision tree algorithms and stratification analyses in the discovery of learners' need for teaching videos [12], even in the prediction of learners' mental state[13], introducing individual variables such as learner's prior knowledge, background, e-learning experience and opening new cognitive and emotional dimensions in the interdisciplinary field of machine learning and smart teaching such as learners' inner needs and motives that affect the working of personality system in general.

All the above-mentioned research trends confirm the value of emotional and cognitive data in smart teaching. Yet, a narrow gap exists between EDM and learning analytics: decision tree models should be built from teachers perspectives to provide guidance to the selection of smart teaching tools and the design of smart teaching and learning activities adaptive to learners' cognitive and emotional needs. However, given the inadequate exploration of the multi-modality of learners' emotional and cognitive data, the existing research paradigms related to learning disability diagnosis are still immature, leaving promising space for the construction of the affective decision tree model with the primary function of learning diagnosis and learning engagement prediction.

The proposed decision tree model in the present research aims to theorize an affective-feedback and affective-assistance system in smart teaching decision-making through the measurement of learners' emotional response to different online learning activities in the PANAS test, to enhance teachers' detection and recognition of learners' affective states in smart learning, optimizing the formulation and implementation of smart teaching strategies. The framework of this research encompasses the following constituents: 1) The general introduction to the empirical design of the PANAS test; 2) The crucial procedures in the construction of the affective decision tree via C5.0 algorithm; 3) The verification of the affective decision tree model in SPSS 26.0; 4) Future research orientations on the application of the affective decision tree in smart teaching, which is shown in Fig. 1:

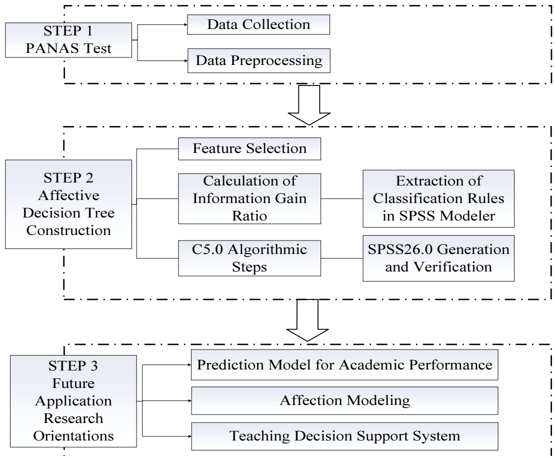


Figure 1. Research flowchart

3. Empirical research into the affective decision tree

3.1 Participants

This research recruited 558 junior college students from a university in Xi'an in 2021. The major of participants covers English and E-commerce. 490 of the participants enrolled in the *Psychology of Foreign Language Teaching* course the author teaches at the university. Supposedly, given their rich experience in blended learning and the prior knowledge in Learning Psychology, the participants have been well acquainted with all the online learning activities and are aware of the correlation between affect, motivation and learning efficiency.

3.2 Data collection

The data in this article is derived from a variant of the PANAS in the form of an on-line survey. The PANAS scales originally devised by David Watson and Lee Anna Clark in the 1980s[14] investigate into the positive and negative mood and feelings the undergraduates experienced in the specified duration in the past. The PANAS scale encompasses descriptors arrayed along two dimensions of the participants' emotional experience, namely, Positive Affect (PA) and Negative Affect (NA). The subjects were asked to rate on a 5-point Likert scale the extent to which they had experienced the listed positive affect and negative affect during a specified time frame. The points of the scale were labeled very slightly or not at all, a little, moderately, quite a bit, and very much, respectively[15].

The replication of PANAS scale is theoretically accountable since its reliability and credibility have been attested in decades of psychometric experiments. The present research adopts the logic of the discrete emotion analysis model embodied in the original PANAS scale for the appropriate classification of learners' academic affects and the selection of reliable affective descriptors.

The revised PANAS scale capitalizes on the affect priming effects of judgments, according to which ambiguous stimuli with less predefined meaning or prior knowledge for an individual tend to trigger up judgments more prominently influenced by affective states. Following the same logic, the present research makes adaptations to the PANAS scale with the addition of the names of 17 online learning activities as stimuli to trigger up learners' academic affects. Theoretically, participants' spontaneous emotional states or traits can be induced through the affective judgments in which they should rate on a 5-point Likert scale the extent to which they associate the listed positive and negative affects with specified online learning activity.

The affective states may be extended to other objects related or unrelated to those stimuli, which is the affect infusion process that operate unconsciously and modulate an individual's cognitive and behavioral responses. Accordingly, the assessment of participants intuitive affective judgments of the correlation between the names of online learning activities and the potential emotions those activities may engender in the course of smart learning is of predictive value since the scores of the test are indicator of participants' preference for the smart learning activities. Similarly, the scores of the test are of high diagnostic value to the teachers since learner's preference is certainly the index of the teaching effect which may in turn influence their pedagogical design.

There are altogether 17 names of the prevalent online learning activities and 8 positive affect descriptors and 8 negative affect descriptors for each category of online learning activities. The items in the affect descriptors form a continuum with the gradation from positive affect to negative affect. The online activities are Audio-Video Learning, Chapter Learning, Online Discussion, Online Homework, PBL Project, Sign-in, Random Selection, Online Test, In-Class Practice, Quick Response, Questionnaire, Voting, Rating, Group Chat, Whiteboard, Live Broadcast, Synchronous Classroom. The 16 items in the affect descriptors are listed in Table 1:

Table 1. Positive and negative affects

Positive Affect	Excited	Happy	Expectative	Contented	Active	Energetic	focused	Relaxed
Negative Affect	Bored	Ashamed	Disappointed	Depressed	Anxious	Nervous	Afraid	Angry

3.3 Data preprocessing

In the original data collected from the empirical survey, the categorical data concerning nominal type attributes such as online learning activity and emotion type is mingled with interval data concerning the numeric type attribute such as learners' rating of affective judgment. For the convenience of data mining process, the categorical data has undergone the data conversion process in SPSS 26.0 for further analysis, which splits the data into two nominal variables: the above-mentioned online learning activity types and the emotion types. The interval data recording learners' rating of the affective experience in the online learning activities is retained. In a 1-5 scale, an affective rating score greater than or equal to 3 is considered a positive index of a specific affective experience. To establish an decision tree model, part of the data has been segmented to form the training set, as is shown in Table 2:

Table 2. Training set

Student Number	Online Learning Activity	Excited	Happy	Expectative	Contented	Active	Energetic	Focused	Relaxed	Bored	Ashamed	Disappointed	Depressed	Anxious	Nervous	Afraid	Angry	Total Score
1	Audio Video Learning	4.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	4.00	3.00	3.00	3.00	3.00	3.00	3.00	2.00	49
2	Audio Video Learning	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	4.00	4.00	4.00	4.00	52
3	Audio Video Learning	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	48
4	Audio Video Learning	1.00	1.00	1.00	1.00	2.00	2.00	3.00	1.00	3.00	3.00	3.00	5.00	5.00	5.00	5.00	1.00	42
5	Audio Video Learning	2.00	2.00	4.00	4.00	4.00	2.00	2.00	5.00	3.00	2.00	3.00	1.00	1.00	1.00	1.00	1.00	38
6	Audio Video Learning	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	2.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	49
7	Audio Video Learning	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	3.00	2.00	2.00	3.00	1.00	1.00	1.00	1.00	22
8	Audio Video Learning	3.00	4.00	2.00	2.00	2.00	2.00	3.00	4.00	2.00	2.00	3.00	4.00	3.00	2.00	3.00	4.00	45
9	Audio Video Learning	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	80
10	Audio Video Learning	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	48
11	Audio Video Learning	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	48
12	Audio Video Learning	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	5.00	3.00	5.00	5.00	5.00	5.00	5.00	5.00	46
13	Audio Video Learning	1.00	1.00	1.00	1.00	1.00	1.00	1.00	2.00	3.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	19
14	Audio Video Learning	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	48
15	Audio Video Learning	4.00	3.00	2.00	1.00	4.00	4.00	4.00	4.00	1.00	1.00	1.00	1.00	3.00	3.00	3.00	3.00	42
16	Audio Video Learning	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	1.00	2.00	2.00	1.00	1.00	2.00	1.00	2.00	36
17	Audio Video Learning	4.00	5.00	5.00	4.00	5.00	4.00	5.00	5.00	4.00	5.00	5.00	4.00	5.00	5.00	5.00	5.00	75
18	Audio Video Learning	3.00	4.00	2.00	3.00	2.00	3.00	2.00	4.00	2.00	3.00	1.00	1.00	1.00	1.00	1.00	1.00	34
19	Audio Video Learning	5.00	5.00	5.00	5.00	5.00	5.00	5.00	5.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	48
20	Audio Video Learning	2.00	2.00	3.00	4.00	4.00	4.00	4.00	4.00	3.00	3.00	3.00	4.00	3.00	3.00	4.00	3.00	53
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...		

3.3 Decision tree construction

- Feature selection

In the training set with the size of 100 subjects, the frequency of each subject's positive affective rating above 3 has been counted to split the data and start the split search. The split search begins by selecting certain input affective variable to be the split. Two subsets are created in the splitting process: the left branch includes all the cases less than the split value, the right branch contains all the cases greater than the split value.

Take the affective variable EXCITED in Table 3 as example, the following procedures have been executed to determine if it is a good split:

Table 3. Descriptive statistics of the affective variable EXCITED

Ratings	Frequency	Percent	Valid Percent	Cumulative Percent
1	15	15.0	15.0	15.0
2	14	14.0	14.0	29.0
3	33	33.0	33.0	62.0
4	23	23.0	23.0	85.0
5	15	15.0	15.0	100.0
Total	100	100.0	100.0	

(1) Running the Pearson's Chi-square statistic quantification of the positive counts in the EXCITED column in SPSS 26.0. The asymptotic significance is 0.010, suggesting the significant difference, which makes the variable a good split.

(2) Calculating the information gain of the affective attribute EXCITED on the training set S, namely, g(E, S). Decision tree learning use information gain in the selection of attributes. The calculation process is shown in Formula (1), (2) and (3):

$$g(E,S) = H(S) - H(E|S) \tag{1}$$

$$H(S) = -\sum_{k=1}^{k} \frac{|C_k|}{|S|} \log_2 \frac{|C_k|}{|S|} \tag{2}$$

$$H(E|S) = \sum_{i=1}^{n} \frac{|S_i|}{|S|} H(S_i) = -\sum_{i=1}^{n} \frac{|S_i|}{|S|} \sum_{k=1}^{K} \frac{|S_{ik}|}{|S|} \log_2 \frac{|S_{ik}|}{|S_i|} \tag{3}$$

In Formula (1) (2) and (3), S stands for a given training set and E stands for a given attribute, accordingly, H(S) is the empirical entropy that represents the classification uncertainty of training set S. Whereas, H(E|S) designates the classification uncertainty of training set S under the given condition of attribute A. The difference between the two is the information gain which denotes the degree of reduction in the classification uncertainty of training set S on account of attribute A. The calculation process is to determine the attribute variable with the maximum classification ability or entropy. In formula (2) , formula (3) and formula (4), |S| stands for the sample size of the training set, in this case, the sample size is 100 and C stands for the categories in attribute E. In this research, two categories are distinguished in attribute E, the positive and negative category respectively. |Ck| stands for the sample size belonging to category Ck, K=2. Since this research adopts a 1-5 rating scale for each emotional type, n=5, which means training set S will be divided into 5 subsets based on its value.

In formula (3), |Si| stands for the sample size of each subset. |Sik| stands for the sample set belonging to a specific Ck, K=2.

Table 4. The input values of H(S|E)

Sik / **Si**	**Si1**	**Si2**
S1	15	0
S2	14	0
S3	0	33
S4	0	23
S5	0	15

Table 4 shows the possible input values of all the letter symbols representing different affective categories and affective ratings in formula (3) and formula (4), which are drawn from the frequency and percentage of descriptive statistics in Table 3: taking 3 as the dividing value, Si1 equals the positive counts and Si2 equals the negative counts belonging to each subset of the training set. In the calculation of H(S|E), the input values listed in Table 4 add up to 0, making empirical entropy H(S) the only valid value to determine information load of the attribute E.

Accordingly, the expected information value for the classification of training set is shown is Formula (4):

$$\begin{aligned} g(E,S) &= -\sum_{\mathrm{k}=2}^{\mathrm{k}} \frac{|C_k|}{|S|}\log_2\frac{|C_k|}{|S|} - [-\sum_{i=1}^{n}\frac{|S_i|}{|S|}\sum_{k=2}^{K}\frac{|S_{ik}|}{|S|}\log_2\frac{|S_{ik}|}{|S_i|}] \\ &= -\frac{71}{100}\log_2\frac{71}{100} - \frac{29}{100}\log_2\frac{29}{100} - 0 \\ &= 0.35 + 0.52 \\ &= 0.87 \end{aligned} \tag{4}$$

The final result of calculation is 0.87, approaching 1, suggesting the information gain of the attribute is the largest, so in the generation of the decision tree, the affective variables are used to construct the internal nodes.

- C5.0 decision tree algorithm steps

The construction of the affective decision tree involves the grouping and matching of multiple type variables designating negative and positive affective categories with optimal online learning activities on the basis of the numeric variables designating learners' negative and positive ratings. Accordingly, C5.0 algorithm is followed to accommodate the multi-branch generation process. In C5.0, the Information Gain Ratio is the standard for the selection of optimal grouping variable as the best binary splitting points. In the generation of the affective decision tree, the input variables are the numeric affective ratings, the output variables are the different affective and the online learning activity types. Taking the probability of each negative or positive affective ratings P as input variable, the calculation of the entropy of the affective variable Excited is shown in Formula (5):

$$
\begin{aligned}
Ent(A_{Excited}) &= -\sum_{Excited} P(A_{Excited})\log_2 P(A_{Excited}) = -\frac{1299}{9486}\log_2(\frac{1299}{9486}) - \frac{1391}{9486}\log_2(\frac{1391}{9486}) \\
&-\frac{3436}{9486}\log_2(\frac{3436}{9486}) - \frac{1649}{9486}\log_2(\frac{1649}{9486}) - \frac{1711}{9486}\log_2(\frac{1711}{9486}) \\
&= 0.394 + 0.407 + 0.530 + 0.440 + 0.446 \\
&= 2.217
\end{aligned} \tag{5}
$$

The calculation of the entropy of a certain affective variable with respect to the input numeric variable is shown in Formula (6):

$$Ent(A_i|R_n) = \sum_n P(R_n)(-\sum_i P(A_i|R_n)\log_2 P(A_i|R_n)) \tag{6}$$

In formula (6), R with the subscript n designates the totality of all the affective ratings, P stands for the affective ratings pertaining to a certain affective variable A_i. For a given affective variable A, all the affective ratings Rn made by subjects fall into two types, the positive, which indicates the strong prominence of an emotion, and the negative, which indicates the weak or undetectable influence of an emotion, which is captured by the subscript i.

As is shown in Table 5, the sum of all the affective ratings made in this research by subjects is 9486, which is further divided along the positive and negative axis by the rating value 3. Consequently, the affective ratings below 3 will be counted as negative, labeled N in Table 5 and the affective ratings that equals 3 or is greater than 3 will be counted as positive, labeled P in Table 5 correspondingly.

Table 5. The descriptive statistics of affective variables

Affective Variable	**Ratings**				
	1	**2**	**3**	**4**	**5**
Category	**N**	**N**	**P**	**P**	**P**
Excited	1299	1391	3436	1649	1711
Happy	1356	1439	3506	1557	1628
Expectative	1319	1388	3535	1561	1683
Contented	1375	1336	3513	1584	1678
Active	1281	1391	3519	1642	1653
Energetic	1318	1424	3500	1627	1617
Focused	1180	1422	3520	1664	1700
Relaxed	1411	1354	3474	1574	1673
Bored	3900	1749	2526	788	523
Ashamed	4412	1673	2361	640	400
Disappointed	4402	1638	2383	623	440

Depressed	4387	1600	2382	640	477
Anxious	4089	1671	2446	783	497
Nervous	4080	1775	2437	692	502
Afraid	4314	1637	2416	659	460
Angry	4667	1505	2294	588	432

Accordingly, the entropy for the affective variable Excited with respect to the numeric affective ratings is shown in Formula (7)

$$\begin{aligned}
&Ent\,(A_{Excited}\,|R_{Excited}\,) = \sum_{Excited} P(R_n)[-\sum_{Excited} P(A_{Excited}\,|R_n)\log_2 P(A_{Excited}\,|R_n)] \\
&= \frac{6796}{9486}[-\frac{3436}{6796}\log_2(\frac{3436}{6796}) - \frac{1649}{6796}\log_2(\frac{1649}{6796}) - \frac{1711}{6796}\log_2(\frac{1711}{6796})] \\
&+ \frac{2690}{9486}[-\frac{1299}{2690}\log_2(\frac{1299}{2690}) - \frac{1391}{2690}\log_2(\frac{1391}{2690})] \\
&= 0.716\,(0.498 + 0.495 + 0.502) + 0.283\,(0.506 + 0.492) \\
&= 1.070 + 0.282 \\
&= 1.352
\end{aligned} \tag{7}$$

The calculation of the information gain for the affective variable Excited is shown in Formula (8)

$$\begin{aligned}
&Gains\,(A_{Excited}, R_{Excited}) = Ent\,(A_{Excited}) - Ent\,(A_{Excited}\,|R_{Excited}) \\
&= 2.217 - 1.352 \\
&= 0.865
\end{aligned} \tag{8}$$

The C5.0 algorithm takes information gain ratio instead of the information gain in the selection of the optimal grouping variable to minimize the interference of type values, therefore, the information gain ratio for the affective variable Excited is shown in Formula (9):

$$\begin{aligned}
&GainsR\,(A_{Excited}, R_{Excited}) = Gains\,(A_{Excited}, R_{Excited}) / Ent\,(A_{Excited}) \\
&= 0.865 / 2.217 \\
&= 0.39
\end{aligned} \tag{9}$$

Following the same algorithmic procedure, table 6 shows the information gain ratio of the other affective variables.

Table 6. The information gain ratio of affective variables

Affective Variable	Information Gain Ratio
Excited	0.39
Happy	0.39
Expectative	0.39
Contented	0.39
Active	0.39
Energetic	0.39
Focused	0.38
Relaxed	0.39
Bored	0.48
Ashamed	0.49
Disappointed	0.49

Depressed	0.49
Anxious	0.48
Nervous	0.48
Afraid	0.49
Angry	0.50

Still, the marginal difference between the information gain ratio of the positive affective variables and that between the negative affective variables respectively makes it difficult to determine the split point, therefore, a C5.0 model is constructed in SPSS modeler to predict the actual importance of those affective variables in tree construction, with online learning activity as the target and the affective variables as the input to highlight the correlation between learners' affective state and teachers' selection of online learning activities. Figure 2 shows the results of prediction:

Figure 2: Predictor importance

From the modeling result, it is found that positive affects like expectation, happiness, excitement and satisfaction are significant influencing factors in learners' selection of smart learning activities. Interestingly, negative affects like nervousness, anxiety, fear or shame also take precedence over positive affects such as attentiveness, relaxation and activeness, which entails that negative affective factors may have some cross influence on the affective priming effect of those positive affective factors.

Therefore, feedback on learners' negative affects is very important in designing the smart teaching and learning activities since negative affects may in a sense override the affective priming effect of positive affects. It is recommended that educators be more sensitive and responsive to learners' negative affects.

- Classification rule construction

After sorting out all the affective variables according to the information gain ratio, the different layers in the architecture of the classification tree can be generated prior to C5.0 modeling in SPSS 26.0, as is shown in figure 3. The architecture mirrors the putative decision-making process guided by the affective decision tree generated in C 5.0 modeling, as is shown in figure 4, with the depth of the tree determined by the value of different affective variables.

The nodes of the of decision tree are affective variables related to learners affective states and affective experience, which affects the decision-making in foreign language teaching. These attribute nodes are designated by rhombuses arranged hierarchically to indicate the different degree of affective priming effect and the resultant classification of learners, judgment of learners' mental states and the selection of teaching and learning activities. Similarly, the different attribute values of affective variables designated by circles constitute the affect-specific internal or test nodes.

In the tree diagram, the test nodes represents two categories of the affective rating, P for positive, with the rating value equaling or greater than 3, and N for negative, with the rating value below 3. The different categories of learning activities designated by rectangles constitute the terminal or decision nodes, they correspond to the personalized or customized teaching methods designed for the strategical and psychological intervention targeted at different affective states in the process of smart learning.

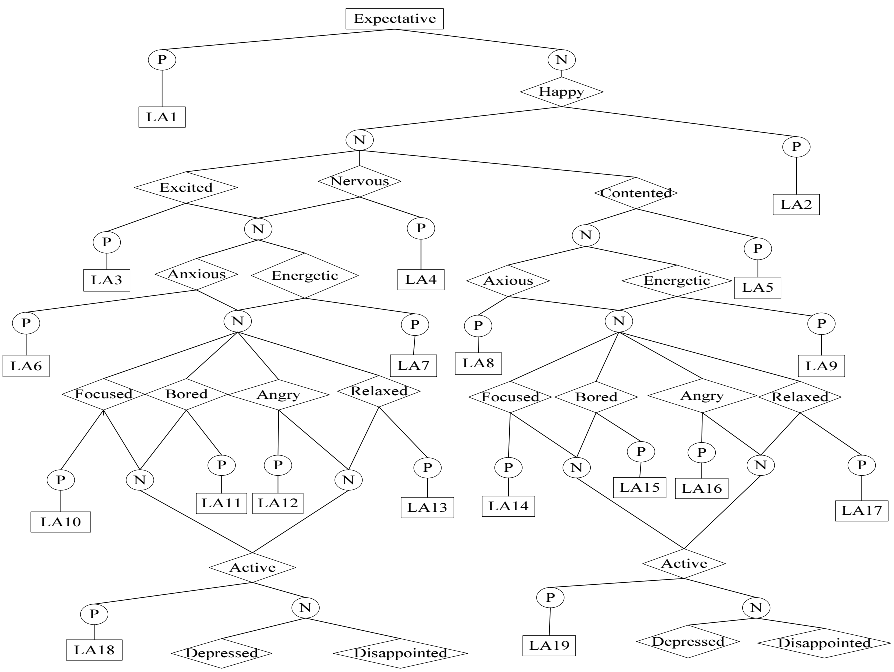


Figure 3. The architecture of affective decision tree

The tree is constructed on the basis of a set of IF···THEN classification rules which determine the path from the root node to each leaf node. The conjuncts of the affective attributes and relevant attribute values formed along a path from the root node constitute the IF part, and the categories of learning activities marked by the leaf nodes constitute the THEN part of the rule, that is, the conclusion of the rule. In C5.0 modeling, the rule determines teachers' pedagogical decision in the selection of online learning activities, taking the online learning activity audio video learning for example, 47 rules are generated in modeling, two of which are listed below for illustration:

Rule 1 for Audio Video Learning:

If Excited <= 2 and Relaxed > 1 and Disappointed > 1 and Disappointed <= 2 and Nervous > 1 and Nervous <= 2 and Afraid > 1 and Angry <= 1 and Student Number > 54 and Student Number <= 402 then Audio Video Learning.

Rule 2 for Audio Video Learning:

If Energetic > 1 and Bored > 1 and Disappointed <= 1 and Anxious <= 1 and Nervous <= 1 and Afraid <= 1 and Angry > 1 then Audio Video Learning

Generally, the rules embody the intricate correlation between the affective mechanism and teachers' pedagogical decision. For instance, teachers should carefully weigh and prioritize the attribute value of affective variables with relatively high information gain ratio, such as Excited, Relaxed, Energetic and Nervous, Afraid, Anxious, Angry to decide when to start Audio Video Learning. Obviously, the rules suggest when the learners are a little disappointed and bored or nervous, Audio Video Learning prompts the working of positive affects like excitement, relaxation. Noticeably, the rules echoes the previous algorithmic results in that negative affects should be the first affective factors to be carefully weighed in smart teaching design, since they have higher information gain ratio rankings.

In the next phase, learners are grouped along different positive and negative axis in affective identification and modeling to decide if they fit in with certain online learning activity. The test and decision procedures may be executed recursively until teachers arrive at the final diagnostic conclusion of learners' affective state and the optimal teaching strategies.

4. The utilization of affective decision tree in smart teaching

To enhance the efficiency of smart teaching and learning, the affective data should be accurately sorted and utilized in learning analytics to establish the data-driven decision-making mechanism. The affective decision tree model fuses learners' affective experience with decision-making analysis to optimize the smart teaching decisions and effect learners' self-adaptive learning. The functions of affective decision tree in smart teaching are threefold:

(1) Predicting students' academic performance

In the multi-modal data fusion stage of the affective decision tree construction, on the basis of the empirical PANAS test, the value and information gain ratio of affective variables indicate the most influencing affective priming effect. Specifically, the affective variable that has the highest information gain ratio and value tend to influence student performance more significantly. Latest empirical research findings have verified that values of affective variables can be used to construct the Affective Feedback Adaptive Learning System employing the recognition-supported techniques to enhance learners' learning engagement and self-directed learning[16]. Furthermore, affective and cognitive constructs figure prominently in the Cognitive-Affective-Motivation Model of Learning (CAMML) which delineates the updated trilogy-of-the-mind (cognitive, conative, affective) model of intellectual functioning. All in all, academic affects, coupled with self-efficacy, play significant mediating roles in online learning: the higher attribute value of negative affective variables normally indicates low learning engagement, which predicts high drop-out rate, low curriculum attendance and inadequate participation in all the 17 categories of online learning activities listed in the research. Conversely, the higher attribute value of positive affective variables indicates high learning engagement, which predicts high frequency of curriculum attendance, learning task completion and learning activity participation, as well as high quality of homework and high rate of knowledge expansion based on efficient prevision and revision.

If the affective variable is further correlated with specific learning activities in the SPSS correlation analysis, a prediction model of those activities can be established to forecast the completion time, academic result or academic performance by value assignment.

(2) Student affection modeling

From the perspective of affective computing, the affective data collected to construct the affective decision tree in the generation stage can be utilized as the new method of coordinating the smart teaching decision with smart learning strategy since all the affective variables can be used to construct the computable and explainable affective model of the students concerning specified smart learning activities, which provides pedagogical guidance to teachers in their smart teaching design. The affective model constitute three basic elements, namely, personality, mood and intentionality, whose valence correlates with different attribute.

As can be inferred from the research findings, the information gain ratio of affective variables determines its position in the affective decision tree. Presumably, the higher ranking of negative affective variables compared with positive affective variables in the hierarchy of the tree entails the malfunction of learners' personality mechanism in coping with the smart learning tasks. Since personality fundamentally directs an individual's behavioral and psychological inclination, the higher ranking of positive affective variables in the hierarchy suggests learners are more inclined to confront challenging tasks in smart learning.

Likewise, along the bidimensional affective axis, the path of the affective decision tree from the internal nodes of positive affective variable to that of negative affective variable shows learners' change of mood in smart learning from positive to negative. The detection of the mood change is very crucial since mood is temporary and transitory, without proper intervention, a mood may be entrenched into a type of emotion or affect associated with certain learning activities, engendering stereotyped mindsets that may be extremely difficult to reverse.

Psychologically, positive affects induces higher intentionality and willingness than negative affects. Therefore, in the execution of smart teaching tasks, the affective decision tree visualizes learners' intention and willingness through branching or splitting rules which classifies learners' into the willing type, with the largest tree branches of positive affect nodes, and unwilling type, with the largest tree branches of negative affect nodes.

(3) Teaching decision support system

After the generation of affective decision tree, teachers may use the architecture of the decision tree as reference for learners' affective fluctuation related to smart learning task distribution.

Hence, in the modeling of decision tree, the student number can be inputted as the dependent variable and the online learning activity is inputted as the independent variable, the affective variables are influencing variables affecting learners' learning decisions and strategies in smart learning. Taking the student number as the splitting variable, a new tree can be constructed, which shows students preferences over certain online learning activity under the influence of specific negative or positive affect. The tree generated in SPSS 26.0 is demonstrated in Figure 4, with the left branch of the tree signaling the more preferable online learning activities learners tend to select under the influence of the specific positive or negative affect, the right branch signaling the less preferable online learning activities teachers should dismiss in smart teaching design. The pruning of the tree is symbolic of the optimization of teaching decisions in screening and filtering the undesirable teaching and learning activities.

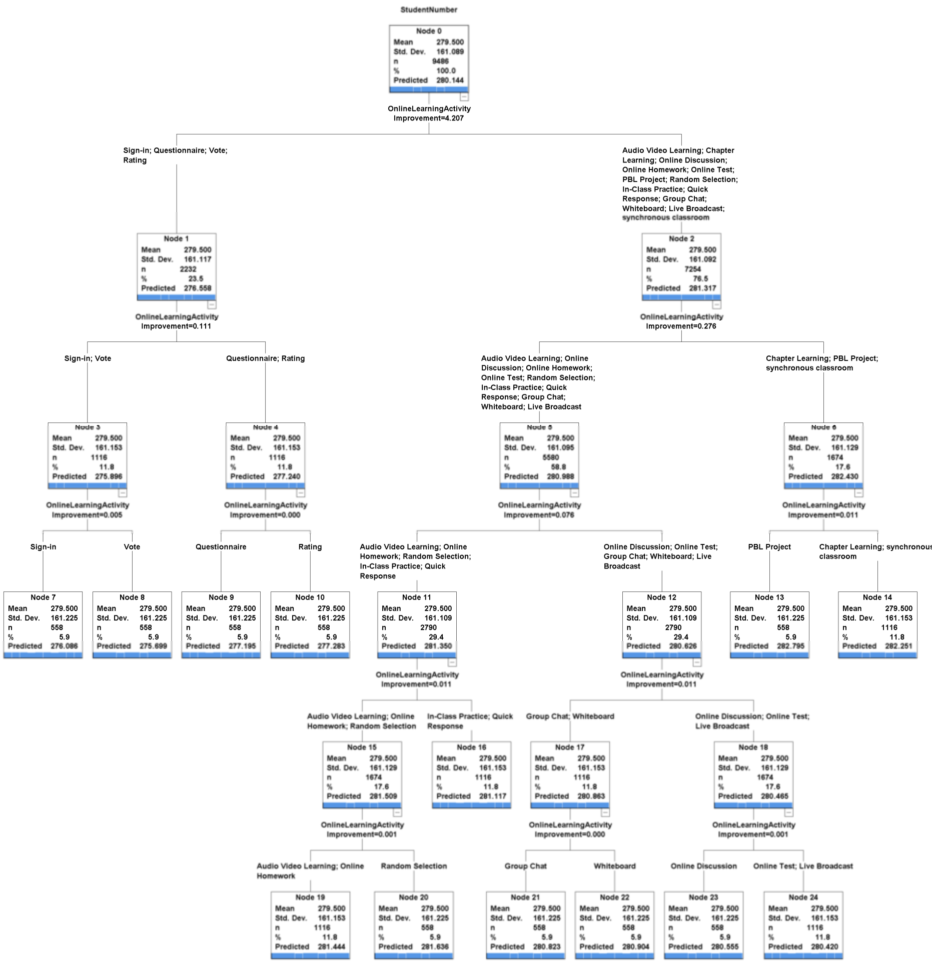


Figure 4. The affective decision tree generated in SPSS 26.0

5. Conclusion

Smart teaching involves constant data-driven decision-making and instructional refurbishment, which makes educational data mining tools like decision tree significant enabler for personalized learning and precise teaching. The commonly-used classification and regression tree models draw on demographic data or procedural and formative data extracted from the learning analytics of the online and offline interaction, learning behavior and progress, test performance can be used in addressing classification and decision problems. However, what qualifies smart teaching as "smart" is the affective understanding and adaption, multimodel-based affective information processing, which means affective data has to be taken into account in smart teaching design since it reveals learners' mentality and mood.

The prospective research presented in this paper sheds light on the new smart teaching paradigm aided by the affective decision tree. The "affective" nature of the tree derives from the affective data collected from the PANAS test. The construction of the affective decision tree leads to the following conclusion:

(1) The architecture of the tree embodies the inner working of learners' affective system which is defined by the cross influence between positive and negative affects. Affective system figures prominently in learners' selection and participation of online learning activities because of emotional infiltration and affective priming effect of positive and negative affects. In smart teaching, affective decision tree can help educators capture the dynamic flow of emotion and establish flexible affective feedback mechanism in making smart teaching decisions.

(2) Theoretically, the affective decision tree can also be used in the construction of prediction models of learners' learning engagement, academic performance, student profiling and learning disability diagnosis, as well as the teaching decision support system.

(3) Hypothetically, the affective decision tree can be utilized in the interdisciplinary field of affective computing and machine learning to materialize the detection, recognition and differentiation of the mentally active, relaxed and stressed learning states in smart teaching. Therefore, the effects of smart teaching can be interpreted in terms of the reciprocal transaction between learners' personality system and affect mechanism and affective decision trees can help teachers achieve affect understanding and cognition in smart learning environment.

References

[1] Xianmin, Yang & Xinquan, Jin. Theories and methodology of smart Learning. Beijing: China Science Publishing & Media Ltd.; 2021. 10p.

[2] Ronghui, Gan&Gaoda, He. Exploring smart education routine for foreign language teaching in 5G era. Technology Enhanced Foreign Language Education. 2021 Apr ; (02):45-51.

[3] Kekang, He. The new development of "learning analytics technology" in China. e-Education Research. 2016 Jul; (07):5-13.

[4] Xiaoqing, Gu et al.. The theoretical framework, development and future prospect of smart education. Journal of East China Normal University (Educational Sciences). 2021 Aug; (08): 20-32.

[5] Hang, Li. Statistical learning method (Edition 2). Beijing: Tsinghua University Press; 2019. 67p.

[6] Kadam S S, Agrawal M V S. . Predictive analysis of placement of students using machine learning algorithm: decision tree. Journal of Data Mining and Management (e-ISSN: 2456-9437). 2022 Dec; 7(1): 1-5.

[7] Dirin, A., & Saballe, C. A. (2022). Machine learning models to predict students' study path selection. International Journal of Interactive Mobile Technologies. 2022 May; 16 (1):16-18.

[8] El Mourabit, I., Jai-Andaloussi, S., Abghour, N. (2022). Educational data mining techniques for detecting undesirable students' behaviors and predicting students' performance: a comparative study. In: Saeed, F., Al-Hadhrami, T., Mohammed, E., Al-Sarem, M., editors. Advances on Smart and Soft Computing. Advances in Intelligent Systems and Computing, vol 1399. Singapore: Springer; 2022. p. 163-170.

[9] Guterres. J. D. D., et.al. Predicting students' performance in basic algorithms programming in an e-Learning environment using decision tree approach. Syntax Literate :Journal Ilmiah Indonesia. 2022 Jan; 7 : 2548-1398.

[10] Shao, Juxin, Qian Gao, and Hui Wang. Online learning behavior feature mining method based on decision tree. Journal of Cases on Information Technology (JCIT). 2022 Feb; 24(5) : 1-15.

[11] Chen, W. K., Chang, J. R., Chen, L. S., & Hsu, R. Y. (2022). Using refined kano model and decision trees to discover learners' needs for teaching videos. Multimedia Tools and Applications. 2022 Feb; 81(6) :8317-8347.

[12] Ahmed, N., Rifat-Ibn-Alam, M., Akib, G. A., Shefat, S. N., & Nandi, D.. An extensive analysis on computing students' academic performance in online environment using decision tree. Turkish Journal of Computer and Mathematics Education (TURCOMAT). 2022 Jan; 13(1), 149-163.

[13] Bo Q., Gao X. Mental state prediction of college students based on decision tree. In: Shi X., Bohács G., Ma Y., Gong D., Shang X., editors. LISS 2021. Lecture Notes in Operations Research. Springer: Singapore; 2022. p. 149-163.

[14] D. Watson, L. A. Clark, and A. Tellegen. Development and validation of brief measures of positive and negative affect: the PANAS scales. Journal of Personality and Social Psychology. 1988 Jun; 54(6):1063-1070.

[15] J. R. Crawford, and J. D. Henry. The positive and negative affect schedule (PANAS): construct validity, measurement properties and normative data in a large non-clinical sample. British Journal of Clinical Psychology. 2004 Dec; 43, (3): 245-265.

[16] Liu HL, Wang TH, Lin HK, Lai CF, Huang YM. The influence of affective feedback adaptive learning system on learning engagement and self-directed learning. Frontier in Psychology. 2022 Apr; 27(13):858411.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230168

Undergraduates' Knowledge Attitude and Behavior (KAB) Towards the Disclosure of Personal Data Online in China

Xiaoyu Li[a], Qin An[b], Wilson Cheong Hin Hong[c], Yunfeng Zhang[d], Kimberly Kolletar-Zhu[e], Xiaoshu Xu[e1]

[a] *Law School, Wenzhou University, Wenzhou City, 325035, China*
[b] *International Business School, Chengdu Institute Sichuan International Studies University, Chengdu, 610000, China*
[c] *Centre for Teaching and Learning Enhancement, Macao Institute for Tourism Studies, China*
[d] *Centre for Portuguese Studies, Macao Polytechnic University, Macao 999078, China*
[e] *School of Foreign Studies, Wenzhou University, Wenzhou City, 325035, China*

Abstract. In the midst of the COVID-19 pandemic, the employment and education sectors have shifted significantly toward online platforms. However, the increased reliance on these digital spaces has raised concerns about personal security information. Scholars have taken note of this issue and have explored its implications, with some employing the extended knowledge, attitude, and behavior (KAB) model to investigate the moderating effects of societal education level on the relationship between knowledge and attitude. Hong et al. [1] conducted a study to examine undergraduates' KAB regarding personal data sharing in Chinese higher education institutions during the pandemic. Using a questionnaire, the study recruited 156 participants from three universities in West and East China. Using SPSS 23.0, data analysis revealed a widespread lack of awareness, a positive attitude, and proper behavior among college students regarding online personal information leakage during the pandemic. Notably, disparities were observed in KAB among students of different grades, majors, and genders. Students in their sophomore, junior, and senior years were found to be more concerned than freshmen about the availability of their personal information online; what's more, science majors were more concerned than students of other majors. There appear to be significant gender differences in personal information sharing, ie., males are more concerned about the security of personal information online than females. Through this study, we aim to emphasize that college students' awareness of personal information protection needs to be improved and suggest that university administrators and policymakers increase information security training. The findings of this study contribute to the theoretical and practical efforts to improve information security in higher education. Future studies should broaden the survey sample and examine the primary factors that influence college students' KAB of personal information security to ensure the generalization of findings.

[1]Corresponding Author: Xiaoshu Xu, Wenzhou University, China; Email: lisaxu@wzu.edu.cn.
Xiaoshu Xu received China National Social Science Late-stage Funding Project (21FJKB013). The questionnaire data used to support the findings of this study have been deposited in the figshare repository (10.6084/m9.figshare.21994346). The authors declare that they have no conflicts of interest to report regarding the present study.

Keywords. KBA model, disclosure of personal data, survey, higher education, online security

1. Introduction

The novel corona-virus pneumonia has a huge impact on all sectors of medicine, health, the economy, and society. In the field of education, the United Nations' Education Policy Brief for the Period and Beyond of COVID-19, released on August 4, 2020, states that the spread of COVID-19 has affected nearly 1.6 billion students in more than 190 countries and territories worldwide, with 94% of students worldwide affected by the closure of schools and educational institutions. In low- and lower-middle-income countries, the proportion is as high as 99 percent. Online education and online working were quite popular throughout the pandemic period, as were numerous apps, online streaming, telemarketing, and so on. As a result, the issue of personal data security arises. The popularity of online learning and data analysis has made learning analytics an essential component of educational technology [2]. In education, the use of students' personal information is increasing, and students' behavior can even be captured and evaluated.

According to the 2021 College Students' Financial Anti-Fraud Research Report, 46% of undergraduate students have been victims of fraud. College and university information security is inadequate.

There is insufficient daily administration and maintenance, network security work is not prioritized, knowledge of information security protection is low, and weaknesses are not addressed or updated on time, among other reasons, because colleges and universities prioritize building but overlook management. All of these qualities put the institution's information system at risk of security breaches. The information security knowledge, abilities, and attitudes of Chinese undergraduates have not received much research. Large-scale empirical studies and updated research data were lacking in the earlier studies.

This study used the "Star" questionnaire to investigate Chinese college students. The questionnaire consists of three parts. The first part focuses on college students' understanding of online personal information security. The second part examines their attitude towards network personal information security, and the third part examines their behavior in network personal information security. The purpose of this study is to understand the current situation of college students' attitudes, knowledge, and ability in personal information security and to make suggestions on the protection of college students' personal information.

2. Materials and Methods

2.1. KAB Model

The Knowledge Attitude Behavior (KAB) model is based on three interrelated parts of the social psychology model: cognition (knowledge), influence (attitude), and behavior, [3] [4].The knowledge-attitude-behaviour (KAB) model was first proposed by Kruger and Kearney to measure information security awareness. The main proposition of KAB is that users have sufficient information security knowledge and a more positive attitude towards information security, which leads to more positive information security behavior. The theoretical framework underpinning KAB entails a comprehension of the interrelationships among its three constituents. Specifically, KAB posits that the progressive accumulation of knowledge in a relevant domain, such as online security, health, or education, will gradually influence an individual's attitude, subsequently instigating a change in their behavior. Knowledge refers to what is known (declarative), how it is known (procedural), when it is known, and why it is known (conditional) [5], whereas attitude and behavior are defined as belief and perception, respectively [6]. The KAB is a dynamic, interactive model that was originally used in the fields of health and environmental psychology, criminology, climate change, and education and is now applied in network security research.

Specifically, Parsons et al. examined the information security loopholes caused by individuals based on (KAB) [7]. The Personal-perspective-based Information Security Questionnaire (HAIS-Q) was developed in 2014, and it outlined the development of its concept as well as the validity and reliability tests [8]. Knowledge of the policies and procedures was also studied, as was the relationship between attitudes towards policies and procedures and the use of work computers (KAB); further knowledge of policies and procedures in 2015 [9]; attitudes towards policies and procedures; and self-reported behavior, combined with organizational factors. Subsequent empirical studies further demonstrated the effectiveness of HAIS-Q as an effective tool for measuring information security factors [10].

McCormac et al. examined the connection between individual information security awareness and individual differences in characteristics using the HAIS-Q (age, gender, personality, and risk-taking). The HAIS-test-retest Q's reliability and internal consistency were both investigated in 2017 [11]. Individual resiliency, workplace stress, and their ISA (KAB) were all studied in 2018 [12].

Sawaya et al. surveyed 3,500 online users from seven countries using the Security Behavior Intention Scale (SeBIS), testing the effectiveness of common security defenses with a special focus on cultural implications. People from Asian countries, particularly Japan, for example, exhibited less safe behavior [13].

Wahyudiwan et al. investigated the ISA level of MERTHE personnel in 2017 using the (KAB) model's three components and the seven key areas of information security. They concluded that knowledge has a positive impact on attitudes and behavior when it comes to information security [14].

Cain et al. conducted 10 cybersecurity-related questions to perform a knowledge and behavior questionnaire survey on corporate employees in 2018, evaluating the impact of age, gender, criminal background, professional expertise, and cybersecurity training [15].

Wiley et al. conducted a 2019 survey of employees in Australian organizations to examine the connection between cybersecurity awareness, organizational culture, and safety culture [16].

Abanoub Riad et al. investigated Estonian dental students' oral health-related (KAB) to promote oral health and disease prevention [17].

2.2. Online Information Security

The student-centered education model does a lot of student data research to understand and help students' efficiency, assist teachers and improve teaching procedures. While the number of research publications on cybersecurity is rising in general, empirical research on security practices in higher education is critically insufficient [18]. For example, Chandarman and Van Niekerk employed the Theory of Planned Behavior Model (TPB) to measure the CSA level of students in South African private higher education institutions [19].

Lean-Ping and Chien-Fatt used the TPB model to investigate the individual information security self-awareness of students at 11 Malaysian universities. This is based on the 2003 National Institute of Standards and Technology Special Report (NIST SP 800-50) [20].

Kim surveyed undergraduates at a business school in the United States about their understanding and attitudes toward information security [21]. Berki et al. assessed prospective IT workers on their knowledge, concepts, and awareness of cybersecurity while utilizing cloud-based services by looking at current IT students' higher education degree programs and cybersecurity courses in five countries: China, Finland, Greece, Nepal, and the United Kingdom [22].

In the studies conducted by Parsons et al., 1112 undergraduates completed the HAIS-Q and participated in lab-based phishing trials [9] [10]. Higher HAIS-Q scores performed better in phishing studies, indicating that HAIS-Q can predict certain aspects of information security behavior.

Vidakis uses the xAPI library to capture data in a serious game environment, which is compatible with the experience API (xAPI) and implemented in the Unity 3D game engine. Use learning analytics in serious games to simplify data generation and record educationally valuable events [2].

2.3. Disclosure of personal data

The concepts of "personal data," "personal information," and "personal data" are used in the legislation of different countries and regions. The Personal Data Ordinance in Hong Kong, China, defines personal data. Personal data is governed by the 2010 "Law on Personal Data Protection" and the 1995 "Law on Computer Processing of Personal Data Protection." The term "personal data" is directly used in legislation in the United Kingdom and the European Union. The concept of "personal data" is used in China's Civil Code and Personal Information Protection Law.

Personal information refers to any sort of information stored electronically or in other ways that, alone or in conjunction with other information, can identify the identity of a specific natural person or reflect the actions of a specific natural person.

Name, date of birth, ID card number, personal biometric information, address, communication contact information, communication records and contents, account passwords, property information, credit information, whereabouts and traces, accommodation information, health and physiological information, transaction information, and other similar information. Information security technology and personal information security standards specify the method and type of personal information to be determined. Personal information should aid in the identification of a specific natural person in two ways: one, through identification, that is, from information to individual; and two, by the information itself, a special identification of a specific natural person. The second is the association, that is, from the individual to the information, such as the known specific natural person, generated by the specific natural person in his or her activities. Information that conforms to one of the above two circumstances shall be judged to be personal information.

According to the 2016 China Personal Information Security and Privacy Protection Report, more than 70% of individuals consider personal information leakage to be a serious problem.

People who know personal information about them, such as their name or place of employment, have called up to 81% of those surveyed. 53 percent have experienced harassment as a result of exposing personal information when searching or browsing the web.

Furthermore, 36% were harassed or defrauded by marketing after their personal information was compromised, such as when renting, purchasing a home, buying a car, taking an exam, or enrolling in college. According to the 2021 National Internet Users' Satisfaction Survey Report on Internet Security, "nearly 80% of Internet users received sales calls from various agents; more than 60% of Internet users received junk mail; and nearly 60% of Internet users received relevant promotional messages."

In addition to harassment, public Internet users can estimate the risk of personal information leakage in other ways, such as: more than forty percent of netizens think of a great data kill because personal information has been leaked; nearly forty percent of netizens check the default agreement to the service agreement; this allows the application to collect user information, which can lead to personal information leakage; and so on.

The "Research Report on OTT Terminal Data Security and Personal Information Protection," published by the China Citic Institute in 2022, highlighted the issues of data security and personal information leakage caused by terminal applications, as well as the serious phenomenon of SDK collection and processing data. According to the report's objective conclusion, personal information security has remained a difficult issue in China in recent years.

During the COVID-19 pandemic in recent years, hundreds of millions of users collected mobile data on a large scale, especially call data logs and social media reports. In previous pandemic studies, researchers have used CDRs provided by mobile network operators to map people's movements. Back in 2014, the GSM Association issued guidelines on privacy when using mobile phone data in response to the Ebola outbreak. However, the long and widespread nature of the current outbreak has led to the use of big data, which has raised privacy and data protection issues.

The Internet of Things is an emerging technology that generates big data. The Internet of Things has been used in a variety of industries, such as health care, home automation, smart cars, and industrial automation. While the characteristics of the Internet of Things provide us with convenience, they also pose risks [23].

The extensive collection and use of data is the main reason for personal information disclosure. For example, in the retail industry, marketing uses decision-making and business planning to assess customer needs. The customer database is processed, and a comparison of age prediction techniques for "Blessed Friday" shoppers based on machine learning is proposed to determine which age group is more interested in "Blessed Friday" [24].

With the advancement of online social networking, undergraduates are active on variety of social networking platforms. As offline communication becomes more difficult, social media platforms such as Twitter and Weibo have evolved into convergence points for online social mindsets. Highly personal, sensitive, and potentially stigmatizing data is disclosed on social networking sites such as Facebook and Weibo [25].

Weinberger et al. [26] investigated and simulated the differences in attitudes toward online privacy and anonymity among male and female Israeli students in order to better understand their security awareness and behavior toward personal information leakage in an epidemic situation [26]. This study looked at undergraduates' personal privacy literacy from three perspectives: knowledge, attitude, and behavior (KAB). Based on Weinberger et al.'s [26] questionnaire, this study created a questionnaire on undergraduates' knowledge, attitudes, and actions regarding personal information security.

3. Research Design

3.1. Research Questions

- Q1: Are undergraduates knowledgeable about the disclosure of personal data online?
- Q2: What are the undergraduates' attitudes toward the disclosure of personal data online?
- Q3: What is the undergraduates' behavior towards the disclosure of personal data online?

In order to find the answers to these questions, we conducted a questionnaire survey of certain undergraduate students in China on their knowledge, attitude, and behavior regarding online personal information security[2]. We attempt to lay out a thorough framework of beliefs, values, and actions. Three components make up the questionnaire. The first component focuses on undergraduates' understanding of online personal information security; the second component examines their attitudes toward network personal information security; and the third component examines their conduct with regard to network personal information security.

The rest of this article is as follows:

- First, we'll go over our research methodology.
- Second, we will present the findings and analysis.
- Third, we present our findings and conclusions.
- Finally, we summarize the study's limitations.

3.2. Research Instruments

First, scales from previous research were consulted during the questionnaire design process of this article, and the scale that was consistent with this study was selected. The researchers inquire with information technology experts about their opinions, update and modify the maturity scale to reflect the specific research situation of this paper, and create a preliminary questionnaire. Second, discuss any suggestions made to the initial questionnaire item set, linguistic expression, etc. In response to feedback, the questionnaire has been modified to make it easier for respondents to accurately define the meaning of the term, to reduce the number of items, and to adjust the measuring items in the words. This alters the final form of the questionnaire. The Likert scale was used to assess students, and the Likert 5 subscales ranged from strongly disagreeing (1) to strongly agreeing (5) as the possibility of answers for all items, with only one answer allowed for each item. The setting of the online questionnaire does not allow blank

[2] Appendix A(1).docx

answers. The advantage of this setting is that it does not lose data and is convenient for accurate data analysis. If respondents do not want to answer, they can directly give up. It only takes about ten minutes to complete the entire questionnaire, and most respondents are willing to complete it all. Third, this research conducts an empirical investigation based on a questionnaire of knowledge, attitude, and behavior related to online personal information security. The data were analyzed using SPSS 23.0.

4. Results and Discussion

In the 2021-2022 academic year, this study is based on a survey of undergraduates from five universities in China. A total of 288 questionnaires were collected in this survey, excluding those that took less than three minutes. A total of 156 valid responses were collected in China. There were 22 students in year one, 114 in year two, 17 in year three, and 3 in year four. 110 were in liberal arts, 10 in engineering, and 36 in natural science. There were 41 males and 114 females. The descriptive information of the participants is shown in Table 1.

First, all undergraduates were screened to inform respondents about the background of the research project. And respondents have a background in higher education , often use the network, and have experienced online learning. Work participants contacted alumni in person, and two English educators invited undergraduates to fill out questionnaires by email and on-site. Senior students, because of the curriculum, busy internships and learning, it is difficult to find senior students. However, the grade and subject distribution of invited students in the sample is uneven.

Table 1. Data description

Data Description	Value Label	N
Year	Freshman Year 1	22
	Sophomore Year 2	114
	Junior Year 3	17
	Senior Year 4	3
Major	Liberal Arts	110
	Engineering	10
	Science	36
Gender	Male	41
	Female	114

4.1. Scale validity

Using Exploratory Factor Analysis (EFA) with Principle Component Analysis and Promax rotation provided by SPSS 23.0, Bartlett's Test of Sphericity was found to be significant ($p<.001$), meeting the assumption of correlations among question items [27]. Kaiser-Meyer-Olkin was at .81, suggesting the sample size is adequate and there are latent variables within the scale [28]. Hence, both criteria for EFA were met.

With Eigenvalue set at 1, EFA indicates that there are 12 variables. The first variable explained 23.71% of total variance, and the second explained 9.85% (see Table 2 for variables and variance percentage).

Table 2. The percentage variance explained by the variables

Variables	Eigenvalues	% of Variance	Cumulative %	Rotation Sums of Squared Loadings
1	12.09	23.7	23.7	7.53
2	5.02	9.85	33.55	6.99
3	3.89	7.63	41.18	5.99
4	2.76	5.42	46.60	5.61
5	2.44	4.78	51.39	4.69
6	1.66	3.25	54.64	5.38
7	1.56	3.05	57.69	2.58
8	1.51	2.97	60.65	3.52
9	1.44	2.81	63.47	5.30
10	1.22	2.40	65.87	3.12
11	1.11	2.18	68.05	2.55
12	1.03	2.02	70.07	4.17

As suggested in the literature, items that did not obtain a loading of 0.3 were suppressed [29] (p. 692). Then, items that did not have any components above 0.3 loading were removed. Questions that that were found to be valid included *knowledge*: k1, k3, k4, k5, k7, k10, k12, k14, k18, k21, k22, k27; *attitude*: a1, a2, a3, a4, a5, a7, a18, a10, a11, a12, a14, a15, a16, a22, a23, a24, a25, a26, a27; and *behaviour*: b1, b2, b3, b4, b5, b7, b8, b9, b10, b11, b12, b13, b14, b18, b19, b21, b22, b23, b24, and b27 (see table 3 for the factor loading of individual items). As can be seen, there are no satisfying items in component 12, which is normal considering its eigenvalue is only marginally above 1 [30].

Table 3. The factor loading of individual items

Loading	1	2	3	4	5	6	7	8	9	10	11	12
k1	0.698											
k3	0.723											
k4	0.478										0.702	
k5											0.722	
k7	0.702											
k10	0.854											
k12	0.736						0.301					
k14	0.557											
k18	0.694											
k21	0.775											
k22	0.508											0.368
k27	0.420											
a1									0.392			

a2								0.638			
a3								0.649	0.351		
a4								0.797			
a5								0.397			
a7							0.821				
a18							0.708			0.321	
a10		0.816									
a11		0.421		0.372							0.294
a12		0.597									
a14		0.837									
a15		0.886									
a16		0.846									
a22	0.672										0.331
a23	0.534				0.564						
a24	0.890										
a25	0.897										
a26	0.951										
a27	0.825										
b1									0.465		
b2									0.826		
b3			0.899								
b4			0.661								
b5			0.767								
b7			0.614								
b8						0.675					
b9						0.752					
b10				0.708							
b11				0.946							
b12				0.602							
b13				0.500		0.321					
b14				0.431		0.305					
b18				0.787							
b19				0.385							
b21					0.809						
b22					0.837						

b23			0.698
b24		0.353	0.558
b27			0.303

4.2. Validity of the data

Reliability describes the consistency of the measured items within a scale. Cronbach's alpha was employed, and the knowledge dimension was found to be at .905 (see Table 4 for the alpha values and descriptive statistics of the knowledge dimension).

Table 4. Reliability Test Scale: Knowledge

Cronbach's Alpha	Cronbach's Alpha coefficients based on standardized entries		Number of items
.905	.905		12
Item Statistics			
	Mean	Std. Deviation	N
k1	3.23	1.15	156
k3	3.13	1.11	156
k4	2.32	1.03	156
k5	2.16	1.07	156
k7	2.94	1.08	156
k10	2.72	1.14	156
k12	2.79	1.16	156
k14	2.38	1.06	156
k18	2.85	1.09	156
k21	2.74	1.19	156
k22	2.14	1.03	156
k27	1.74	0.90	156

We then analyzed the internal reliability of attitudes toward online personal information leakage. Cronbach's alpha was found to be at .896 (see Table 5 for the Alpha values and descriptive statistics of the attitude dimension).

Table 5. Reliability Test Scale: Attitude

Cronbach's Alpha	Cronbach's Alpha coefficients based on standardized entries		Number of items
.896	.896		19
Item Statistics			
	Mean	Std. Deviation	N
a1	3.13	1.11	156
a2	2.90	1.24	156
a3	2.78	1.12	156
a4	2.86	1.10	156
a5	2.60	1.03	156
a7	2.18	1.02	156
a10	3.54	1.10	156
a11	3.12	1.05	156
a12	3.33	1.09	156
a14	3.53	1.14	156
a15	3.45	1.09	156
a16	3.37	1.05	156
a18	2.76	1.13	156
a22	2.79	1.12	156
a23	3.10	1.20	156

a24	2.56	1.07	156
a25	2.63	1.11	156
a26	2.58	1.08	156
a27	2.42	1.11	156

Finally, Cronbach's alpha of behaviour was at .857 (see Table 6).

Table 6. Reliability Test Scale: Behavior

Cronbach's Alpha	Cronbach's Alpha coefficients based on standardized entries		Number of items
.857	.857		20
	Item Statistics		
	Mean	**Std. Deviation**	**N**
b1	2.94	1.11	156
b2	2.58	1.23	156
b3	2.03	.89	156
b4	2.34	.98	156
b5	2.08	.96	156
b7	1.67	.92	156
b8	3.11	1.21	156
b9	3.24	1.22	156
b10	3.64	1.13	156
b11	3.12	1.17	156
b12	3.32	1.10	156
b13	3.53	1.28	156
b14	3.68	1.27	156
b18	2.39	1.13	156
b19	3.29	1.10	156
b21	3.13	1.16	156
b22	2.70	1.12	156
b23	2.90	1.21	156
b24	2.26	1.08	156
b27	1.94	.97	156

Hence, it can be concluded that both the validity and reliability of the proposed scale was satisfactory.

4.3. Assumption testing for regression analysis

The variables' normality was examined. Because there were more than 50 participants, predetermined tests like Kolmogorov-Smirnov and Shapiro-Wilk were inappropriate. A manual evaluation was used to determine the skewness and kurtosis values. The term "normal distribution" was used to describe samples with skewness between -2 and +2 and kurtosis between -7 and 7. Average knowledge (skewness = -.240, kurtosis = 1.31), attitude (skewness = -.40, kurtosis =-.66), and behavior (skewness = -.17, kurtosis =.36) were all within the acceptable threshold range and were normally distributed.

Table 7. Skewness and Kurtosis

		Statistic	Std. Error
knowledge_mean	The average	2.60	0.06
	Partial degrees	0.24	0.19
	kurtosis	0.42	0.39
attitude_mean	The average	2.93	0.05
	Partial degrees	0.39	0.19
	kurtosis	0.66	0.39
behavior_mean	The average	2.79	0.05

	Partial degrees	0.17	0.19
	kurtosis	1.31	0.39

Figures 1-Figure 3 show that the data are roughly normally distributed.

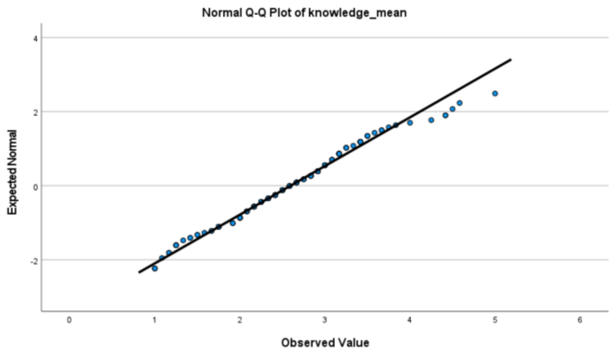


Figure 1. Plot of knowledge mean

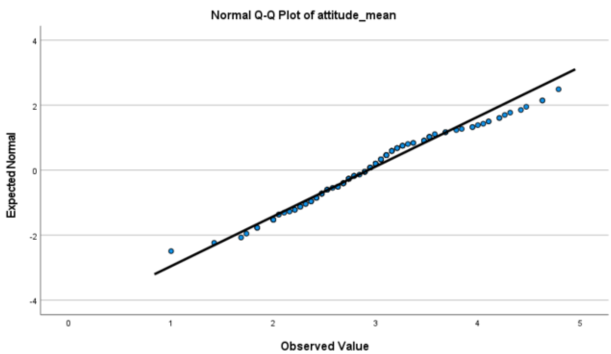


Figure 2. Plot of attitude mean

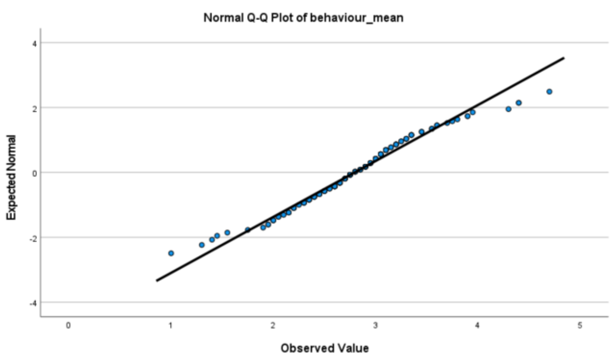


Figure 3. Plot of behavior mean

Figures 1-3 show that the data develops roughly along the normal distribution line in the quantile plot, with some deviations in the tail.

Based on these figures, we can safely assume that this set of data is normally distributed.

Because equal variance is not achieved, Pillai's Trace can only be used for analysis, where the significance test shows that the significance value of gender is 017 [31].

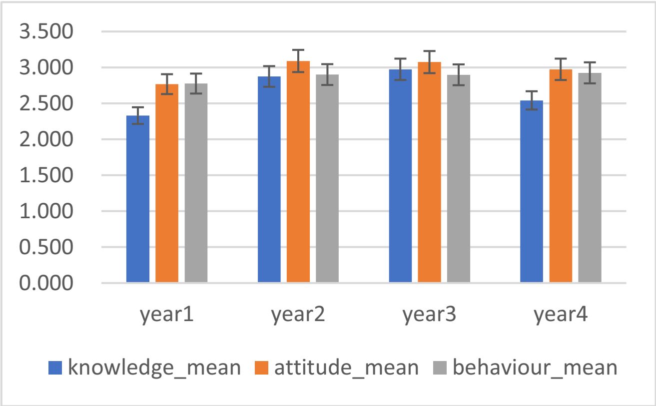


Figure 4. Average values and differences in knowledge, attitude, and behavior of different years.

Figure 4 depicts the average values and differences in knowledge, attitude, and behavior of different grades regarding the disclosure of personal information online. The mean values of the freshman's knowledge, attitude, and behavior are 2.330, 2.768, and 2.776, respectively. Sophomores' mean values for knowledge, attitude, and behavior were 2.875, 3.090, and 2.901, respectively. Juniors' mean knowledge, attitude, and behavior scores were 2.974, 3.075, and 2.898, respectively. Seniors' mean values for knowledge, attitude, and behavior were 2.542, 2.974, and 2.925, respectively. In terms of knowledge, the difference between freshmen and juniors was the largest. In terms of attitude, the difference was greatest between freshmen and sophomores. In terms of behavioral averages, the freshman and senior years differed the most. Less understanding to understanding, less agreement to mostly agreement, and less agreement to mostly agreement are the mean values for knowledge, attitude, and behavior, respectively.

Further investigation reveals that senior students are more concerned about online personal information disclosure than junior students (freshmen). Senior students have a relatively low level of knowledge on this subject, which should be noted. Seniors clearly have more experience with and understanding of online social networking and shopping than juniors.

Furthermore, seniors are about to enter society and will be more concerned with social issues as well as personal rights and interests. More than half of all undergraduates who have been victimized are freshmen, according to law enforcement statistics. Freshmen appear to be less concerned about online personal information disclosure.

Figure 5 depicts the average values and professional differences in the three aspects of knowledge, attitude, and behavior for various majors. In general, the mean values of knowledge, attitude, and behavior in the liberal arts are 2.599, 2.926, and 2.806, respectively. The mean values of knowledge, attitude, and behavior in engineering are 2.442, 2.800, and 2.895, respectively. The mean values of knowledge, attitude, and behavior in science were 3.000, 3.204, and 2.904, respectively. In terms of knowledge, the difference between engineering and science was the largest. In terms of attitude, the difference was greatest between engineering and science. In terms of behavior, the difference was greatest between science and engineering. Less understanding of understanding, less agreement to mostly agreement, and less agreement to mostly agreement are the mean values of knowledge, attitude, and behavior, respectively.

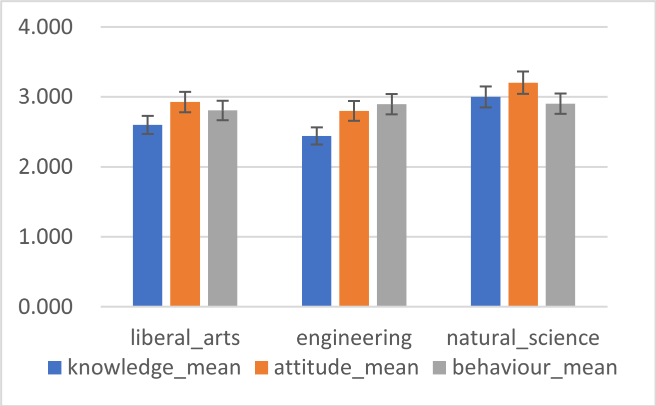


Figure 5. Average values and professional differences of the three aspects of knowledge, attitude, and behavior of different majors

Further investigation revealed that the values of knowledge, attitude, and behavior of undergraduates from all majors regarding personal information disclosure were greater than 2.4. Students majoring in science are more concerned about online personal information disclosure than students majoring in other fields. Engineering students are less concerned about the problem. The gap between genders, the average knowledge, the average attitude, and the average behavior of different genders are shown in Figure 6.

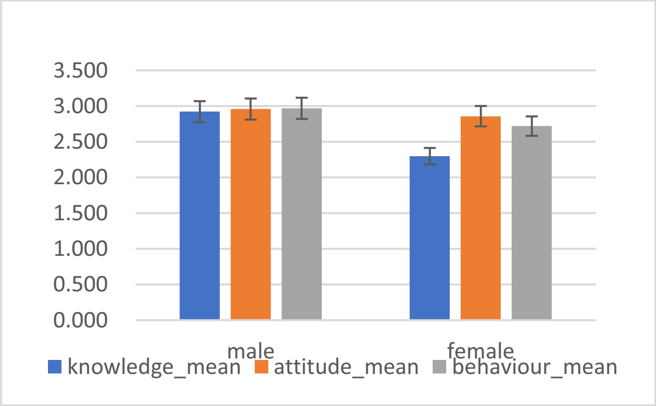


Figure 6. Average values and gender differences of knowledge, attitude, and behavior of different genders

Figure 6 depicts the average values and gender differences in knowledge, attitudes, and behaviors. In general, the mean values of boys' knowledge, attitude, and behavior were 2.922, 2.958, and 2.968, respectively. Girls' mean knowledge, attitude, and behavior scores were 2.298, 2.858, and 2.719, respectively. Less understanding of understanding, less agreement to basic agreement, and less agreement to basic agreement were associated with lower mean values of knowledge, attitude, and behavior, respectively.

Then, a multiple regression analysis was run, which suggests that there are significant differences among the tested variables of grade level, discipline, and gender. However, multivariate analysis reveals that only gender had significant differences (<.001).

On the one hand, for the knowledge, attitude, and behavior of the surveyed personal information network, the average value for a male is higher than that of a female. The difference in knowledge about the disclosure of personal information online was the largest. The conclusion is that undergraduates do not place a high value on the disclosure of personal information online, and male students place a higher value on the security of personal information online than female students.

5. Conclusions

This paper examines the knowledge, attitude, and behavior of Chinese undergraduates' online information disclosure from three aspects. Using quantitative analysis, this paper conducts a comprehensive investigation into the disclosure of Chinese undergraduates' online personal information through the questionnaire research method and uses SPSS 23.0 to conduct correlation analysis and regression analysis on the collected sample data.

The results of the questionnaire survey revealed that undergraduates' knowledge, attitude, and behavior regarding the disclosure of personal information online during the epidemic were not optimistic, with the majority scoring far below the satisfactory score of 4. Schools need to increase courses on personal information protection. Although there is a certain awareness of information security, the overall level of attention is low. Essentially, it was in the stage of "less understanding to understanding, less agreement to basic agreement, and less agreement to basic agreement." Furthermore, gender, grade, and other major factors influence undergraduates' attention to this issue. We investigated the differences in undergraduates' knowledge, attitudes, and behaviors regarding the disclosure of personal information online by grade, major, and gender.

According to the findings, senior students were more concerned about the disclosure of personal information online than junior students (freshmen). Students majoring in science are more concerned than students majoring in other fields about the disclosure of online personal information. Engineering students are less concerned about the issue. Although there are only 10 engineering students in this survey, this conclusion verifies a social survey in 2017. According to a survey conducted by the organizing committee of the 2017 China Media Leaders Conference and the Social Survey Center of Shanghai Jiao Tong University, QQ is the most popular social media platform used by undergraduates, followed by WeChat and Weibo. Engineering students are the most likely to use QQ media. Because engineering students rely too heavily on social networking platforms and have too much faith in information technology, they fail to consider the issue of online personal information disclosure. There is a significant gender difference in the disclosure of personal information on the Internet, with male students paying more attention to the safety of personal information online than female students. Students know about general security threats and protection procedures. However, they did not take sufficient measures to protect their devices or information and did not follow good information security practices.

This finding is consistent with previous research. According to Weinberger et al. [26] women's relatively high level of online privacy self-efficacy (which may be based on their lower level of technological threat awareness) is not matched by their relatively low level of technological online privacy literacy. As a result, they are less capable than men of safeguarding their identity and personal information. On the other hand, men are more aware of technological threats than they are of their online privacy self-efficacy, which, when combined with their relatively high online privacy literacy, enables them to better protect their identity and personal data.

There are few provisions in Chinese law that specifically protect students' personal information, which is one reason for the frequent incidents of students' personal information leakage. Schools also have shortcomings in protecting students' personal information, which include: a lack of awareness of personal information protection, a lack of corresponding protection mechanisms, a low level of network security technology, and insufficient education for students. In light of this issue, the author believes that education and publicity should be prioritized. Schools should offer relevant courses, lectures, forums, and other events to publicize and explain personal information protection issues. Students can spontaneously form relevant associations, create relevant public homepages on Weibo, WeChat, and other social networking sites, and use the network platform to make undergraduates more conveniently and frequently access knowledge about personal information protection, thereby improving their cognition and understanding of the issue. Undergraduates should also focus on increasing their awareness of personal information protection and protecting their rights and interests in personal information. They should, for example, refuse to provide personal information to an information collection organization they do not trust. Avoid using your own real data in general business processing. If you must use real information, you should carefully read the terms of personal information protection and write down the terms of service. Providers should be held accountable for their own violations of the provisions and the method of obtaining compensation. When you discover that your personal information has been compromised, you should immediately contact the appropriate institutions to stop the infringement and negotiate a resolution. Simultaneously, through administrative supervision procedures, you should request that the special regulatory agencies implement the necessary administrative supervision of the infringement agencies and infringements. When administrative actions are ineffective, they should seek legal recourse.

Among them, Article 66 of the People's Republic of China's Personal Information Protection Law specifies administrative liability rules for improper handling of personal information. In conjunction with other provisions, the entire process of handling personal information, such as collection, storage, processing, sharing, transfer, disclosure, and destruction, shall be included in the scope of the regulation. The authority to punish departments responsible for personal data protection in various circumstances has been clarified. Illegal acts are classified into two types: general acts and serious acts. For serious illegal enterprises, the maximum fine limit is less than 50 million renminbi, or less than 5% of the previous year's turnover. While regulating illegal enterprises, relevant responsible personnel are also punished, and senior management of enterprises may be barred from working for an extended period of time. In order to regulate and govern the security of personal information collected from apps and protect consumers' legitimate rights and interests, the China Consumers Association proposed the following in 2018: First, expedite the Personal Information Protection Law legislative process. Second, research the scope and means of putting the real-name system in place. Third, we will increase the level of supervision and inspection. Fourth, we will step up the crackdown even more. Fifth, increase consumer awareness.

6. Limitations and future directions

This paper investigates undergraduates' personal information security literacy in three dimensions: knowledge, attitude, and behavior (KAB). Based on Weinberger et al.'s [26] questionnaire, the study constructed a modified questionnaire regarding undergraduates' knowledge, attitudes, and behaviors toward personal information security. Some limitations need to be addressed in future research: first, a limited number of participants; this study only involved three universities in the East and West regions of China regarding 156 undergraduates' online information disclosure. In future research, more participants from different regions and levels of universities in China could take part in the investigation, which could expand the display and generalization of data. Second, research on other factors that influence undergraduates' knowledge, attitudes, and behavior toward personal information security could be investigated to further research on personal data and Information Security in higher education. Finally, although the survey object of this study is college students, future research can be extended to junior high school students, graduate students, teachers, and so on.

References

[1] Hong WCH, Chi C, Liu J, Zhang YF, Lei VN-L, and Xu XS, "The influence of social education level on cybersecurity awareness and behaviour: a comparative study of university students and working graduates," Education and Information Technologies, Jun. 2022, doi: https://doi.org/10.1007/s10639-022-11121-5.

[2] Vidakis N, Barianos AK, Trampas AM, Papadakis Kalogiannakis, SM and Vassilakis K, "In-Game Raw Data Collection and Visualization in the Context of the 'ThimelEdu' Educational game. In: International Conference on Computer Supported Education," in CCIS Series Book, Heraklion Crete-Greece, May 2019, pp. 629–646.

[3] Eifert G and Craill L, "The Relationship between affect, behaviour, and cognition in behavioural and cognitive treatments of depression and phobic anxiety," Behaviour Change, vol. 6, no. 2, pp. 96–103, 1989, doi: https://doi.org/10.1017/S0813483900007634.

[4] MacKinnon NJ and Hoey J, "Operationalizing the relation between affect and cognition with the somatic transform,"EmotionReview,vol.13,no.3,pp.245–256,Jun.2021,doi: https://doi.org/10.1177/17540739211014946

[5] Schrader PG and Lawless KA, "The knowledge, attitudes, & behaviors approach how to evaluate performance and learning in complex environments," Performance Improvement, vol. 43, no. 9, pp. 8–15, Sep. 2004, doi: https://doi.org/10.1002/pfi.4140430905.

[6] An Q, Hong WCH, Xu XS, Zhang YF, and Kolletar-Zhu K, "How Education Level Influences Internet Security Knowledge, Behaviour, and Attitude: A Comparison among Undergraduates, Postgraduates and Working Graduates," International Journal of Information Security, pp. 1–13, 2022,doi: https://doi.org/10.1007/s10207-022-00637-z

[7] Parsons K, McCormac A, Pattinson MR, Butavicius M, and Jerram C, "An analysis of information security vulnerabilities at three Australian government organisations," in EISMC, 2013, pp. 34–44.

[8] Parsons K, McCormac A, Butavicius M, Pattinson M, and Jerram C, "Determining employee awareness using the Human Aspects of Information Security Questionnaire (HAIS-Q)," Computers & Security, vol. 42, no. 1, pp. 165–176, May 2014, doi: https://doi.org/10.1016/j.cose.2013.12.003.

[9] Parsons K, Young E, Butavicius M, McCormac A, Pattinson MR, and Jerram C, "The Influence of Organizational Information Security Culture on Information Security Decision Making," Journal of Cognitive Engineering and Decision Making, vol. 9, no. 2, pp. 117–129, May 2015, doi: https://doi.org/10.1177/1555343415575152.

[10] Parsons K, Calic D, Pattinson M, Butavicius M, McCormac A, and Zwaans T, "The Human Aspects of Information Security Questionnaire (HAIS-Q): Two further validation studies," Computers & Security, vol. 66, no. 1, pp. 40–51, May 2017, doi: https://doi.org/10.1016/j.cose.2017.01.004.
[11] McCormac A, Zwaans T, Parsons K, Calic D, Butavicius M, and Pattinson M, "Individual differences and information security awareness," Computers in Human Behavior, vol. 69, no. 1, pp. 151–156, Apr. 2017, doi: https://doi.org/10.1016/j.chb.2016.11.065.
[12] McCormac A, Calic D, Parsons K, Butavicius M, Pattinson M, and Lillie M, "The effect of resilience and job stress on information security awareness," Information & Computer Security, vol. 26, no. 3, pp. 277–289, Jul. 2018, doi: https://doi.org/10.1108/ics-03-2018-0032.
[13] Sawaya Y, Sharif M, Christin N, Kubota A, Nakarai A, and Yamada A, "Self-confidence trumps knowledge: A cross-cultural study of security behavior," in ACM SIGCHI Conference on human factors in computing systems, 2017, pp. 2202–2214.
[14] Wahyudiwan DDH, Suchyo YG, and Gandhi A, "Information security awareness level measurement for employee: Case study at Ministry of Research, Technology, and Higher Education," in 3rd International Conference on Science in Information Technology: Theory and Application of IT for Education, 2017, pp. 654–658.
[15] Cain AA, Edwards ME, and Still JD, "An exploratory study of cyber hygiene behaviors and knowledge," Journal of Information Security and Applications, vol. 42, no. 1, pp. 36–45, Oct. 2018, doi: https://doi.org/10.1016/j.jisa.2018.08.002.
[16] Wiley A, McCormac A, and Calic D, "More than the individual: Examining the relationship between culture and Information security awareness," Computers & Security, vol. 88, no. 1, p. 101640, Jan. 2020, doi: https://doi.org/10.1016/j.cose.2019.101640.
[17] Riad A et al., "Estonian Dental Students' Oral Health-Related Knowledge, Attitudes and Behaviours (KAB): National Survey-Based Study," International Journal of Environmental Research and Public Health, vol. 19, no. 3, p. 1908, Feb. 2022, doi: https://doi.org/10.3390/IJERPH19031908.
[18] Ulven JB and Wangen G, "A systematic review of cybersecurity risks in higher education," Future Internet, vol. 13, no. 2, p. 39, Feb. 2021, doi: https://doi.org/10.3390/fi13020039.
[19] Chandarman R and Van Niekerk B, "Students' cybersecurity awareness at a private tertiary educational institution," The African Journal of Information and Communication, vol. 20, no. 1, pp. 133–155, Dec. 2017, doi: https://doi.org/10.23962/10539/23572.
[20] Lean-Ping O and Chien-Fatt C, "Information security awareness: An application of psychological factors - A study in Malaysia," in CCIT, International Conference on Computer, 2014, pp. 98–101.
[21] Kim EB, "Information security awareness status of business college: Undergraduate students," Information Security Journal: A Global Perspective, vol. 22, no. 4, pp. 171–179, Jul. 2013, doi: https://doi.org/10.1080/19393555.2013.828803.
[22] Berki E, Kandel CS, Zhao Y, and Chaudhary SA, "Comparative study of cyber-security knowledge in higher education institutes of five countries," in EDULEARN17 Conference 3rd-5th, 2017.
[23] Shaukat K, Alam TM, Hameed IA, Khan WA, Abbas N, and Luo S, "A Review on Security Challenges in Internet of Things (IoT)," IEEE Xplore, pp. 1–6, Sep. 2021,doi: https://doi.org/10.23919/ICAC50006.2021.9594183.
[24] Shaukat K, Luo S, Abbas N, Mahboob Alam T, Ehtesham Tahir M, and Hameed IA, "An Analysis of Blessed Friday Sale at a Retail Store Using Classification Models," in 2021 The 4th International Conference on Software Engineering and Information Management, Jan. 2021, pp. 193–198. doi: https://doi.org/10.1145/3451471.3451502.
[25] Nosko A, Wood E, and Molema S, "All about me: Disclosure in online social networking profiles: The case of FACEBOOK," Computers in Human Behavior, vol. 26, no. 3, pp. 406–418, 2010.
[26] Weinberger M, Zhitomirsky-Geffet M, and Bouhnik D, "Sex differences in attitudes towards online privacy and anonymity among Israeli students with different technical backgrounds," Information Research, vol. 22, no. 4, pp. 1–23, 2017.
[27] Bartlett MS, "Properties of sufficiency and statistical tests," in Proceedings of the Royal Society of London. Series A - Mathematical and Physical Sciences, May 1937, vol. A-160, no. 901, pp. 268–282. doi: https://royalsocietypublishing.org/doi/10.1098/rspa.1937.0109.
[28] Kaiser HF and Rice J, "Little Jiffy, Mark Iv," Educational and Psychological Measurement, vol. 34, no. 1, pp. 111–117, Apr. 1974, doi: https://doi.org/10.1177/001316447403400115.
[29] Field A, Discovering Statistics Using IBM SPSS Statistics, 5th ed. Los Angeles: Sage Publications, 2018.
[30] Cliff N, "The eigenvalues-greater-than-one rule and the reliability of components.," Psychological Bulletin, vol. 103, no. 2, pp. 276–279, 1988, doi: https://doi.org/10.1037/0033-2909.103.2.276.
[31] Anderson MJ and Walsh DCI, "PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: What null hypothesis are you testing?," Ecological Monographs, vol. 83, no. 4, pp. 557–574, Nov. 013, doi: https://doi.org/10.1890/12-2010.1.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230169

From Social Presence to Virtual Presence: Insights into E-Commerce Consumer Behavior

Jing Deng[1], Rosita Binti Mohd Tajuddin, Benhui Ren, Zijie Chen
College of Creative Arts, University Technology Mara (uitm), 40450 Shah Alam, Malaysia

Abstract. Live streaming has brought new opportunities for e-commerce development. As the most distinctive feature of live e-commerce is the virtual presence. How it creates a more immersive and engaging online shopping experience for consumers is a concern for academics and businesses alike. This paper reviews the relevant literature and analyzes the impact of the evolution from social presence to virtual presence on consumer behavior since the development of e-commerce, summarizing the interactions and main differences between social presence and virtual presence. The research results help e-commerce brands to better understand the differences and maximize their potential in different e-commerce models, so as to gain a deeper understanding of consumer needs and improve positive behaviors such as consumer purchase decisions and loyalty.

Keywords. E-commerce, consumer behavior, electronic brand marketing, virtual presence, social presence.

1. Introduction

The COVID-19 pandemic has affected the world and has influenced the widespread use of social media and digital marketing. Increasingly, more consumers are shifting from traditional online shopping to using social media platforms to discover and purchase products[1] (such as Instagram Shopping, Facebook Live Shopping, Amazon Live, TikTok, etc.). These live-streaming e-commerce consumers are driving e-commerce to evolve from simple online shopping to a more immersive and interactive experience[2].

Previously, scholars have mostly studied live-streaming consumer behavior from the perspectives of anchor traits, social influence, entertainment, cultural factors and Technology Acceptance Model (TAM). The most significant feature of live e-commerce (virtual presence) is overlooked, and it can bring to live e-commerce characteristics such as strong interactivity and real-time, thus distinguishing it from traditional e-commerce, which is valued by e-commerce brands and consumers [3].

Through the study, it is found that enhancing the social presence and virtual presence of consumers in live shopping can increase the inherent social needs of emotional belonging, communication, exchange and identification between anchors and consumers, and between consumers and consumers, thus deepening consumers' awareness of brands,

[1] Corresponding Author, Jing Deng, College of Creative Arts, University Technology Mara (uitm), 40450 Shah Alam, Malaysia. E-mail:284528074@qq.com

anchors and products, and increasing the "stickiness" and loyalty between them and brands.

2. The development process of E-commerce

In recent years, e-commerce has made significant developments worldwide, with more and more consumers turning to online shopping as a convenient way to purchase goods and services[4]. The rapid development of the internet and widespread use of smartphones, along with the rise of social media and digital marketing, have promoted the rapid evolution from traditional e-commerce to live-streaming e-commerce[5]. Live streaming e-commerce facilitates real-time interaction between online retailers and consumers, creating more attractive and personalized shopping experiences. The evolution of development from traditional e-commerce to live-streaming e-commerce can be traced back to several periods:

- Text-based e-commerce (1990s): The earliest online shopping sites such as Amazon and eBay were mainly text-based, providing product descriptions and prices for customers to browse and purchase[6].
- Visual-based e-commerce (early 2000s): The introduction of visual content such as images and videos made e-commerce more attractive and user-friendly, enabling consumers to better visualize products and make more informed purchasing decisions [7].
- Mobile e-commerce (post-2000s): The rise of mobile devices and mobile internet usage created new channels for e-commerce, allowing consumers to shop anytime, anywhere [8].
- Real-time streaming e-commerce (since 2016): With the rise of platforms like Taobao Live and Douyin (the Chinese version of TikTok), live-streaming e-commerce began to gain popularity in China around 2016, and spread to other parts of the world. Integrating real-time video streams into e-commerce, providing real-time interaction and engagement between hosts and consumers, has become an important trend in the e-commerce industry[9].

In summary, live-streaming e-commerce builds upon the advantages of previous periods of e-commerce and adds new levels of interactivity and personalization to the shopping experience. It allows brands and retailers to create engaging entertainment activities while providing consumers with a more authentic and trustworthy shopping experience. Therefore, live-streaming commerce may continue to grow and expand in the coming years.

3. The process of the development of social presence and virtual presence

Social presence and virtual presence are two related concepts. Social presence refers to the degree to which individuals feel a sense of connection and interpersonal relationships with others in communication environments, with a focus on the social aspects of online interaction. In contrast, virtual presence refers to the degree to which individuals feel a sense of physical presence in virtual environments, with an emphasis on the sensory and perceptual aspects of the virtual environment.

- Early virtual reality systems, such as flight simulators and early gaming systems, provided limited social presence and virtual presence. The concept of social presence was first introduced by John Short (1976)[10], defined as "the degree of salience of the other person in the interaction and the consequent salience of the interpersonal relationship." This concept was initially developed in computer-mediated communication environments and has since been applied to other forms of communication technology and settings, including social media, online games, and virtual reality.
- In the early days of the internet in the 1990s, social presence in e-commerce was often limited to text-based chat systems that provided only basic levels of interaction but lacked the richness and nuances of face-to-face communication. The development of social commerce marked the period when social presence became most prominent in e-commerce. The integration of social media and e-commerce created a more social and interactive shopping experience, enabling consumers to share and recommend products on social networks[11].
- As researchers began to explore the potential of virtual reality and other immersive technologies, virtual reality technology first emerged as a way to simulate real-world experiences in the late 1990s and early 2000s. Biocca (1992) [12]was one of the pioneers in developing the concept of virtual presence. He defined virtual presence as the feeling of being present in a virtual environment and interacting with it in a natural and intuitive way.
- In recent years, technological advancements, including high-speed internet, mobile devices, and widespread use of social media, have led to the emergence of new forms of virtual presence in e-commerce, such as virtual showrooms, 360-degree product views, augmented reality experiences, and live streaming e-commerce. Live streaming e-commerce, in particular, has become increasingly popular in many parts of the world, especially in Asia. This approach combines virtual presence with social presence, allowing consumers to interact with hosts and other shoppers in real-time, ask questions, and receive personalized recommendations[13]. Live streaming e-commerce is believed to drive sales growth for many businesses, especially in the fashion, beauty, and food industries.

In general, the development of virtual presence in e-commerce has changed the way consumers shop online and interact with businesses, providing a more immersive and engaging experience and helping to drive the growth of many companies. With technology constantly advancing, we are likely to see even more innovative uses of virtual presence in e-commerce in the coming years.

4. The relationship between social presence and virtual presence in e-commerce.

Although many researchers believe that virtual presence includes social presence, this article aims to provide a clearer analysis and summary of the relationship (Table 1) and main differences (Table 2) between these two concepts.

4.1. The correlation between social presence and virtual presence

Social presence and virtual presence are interconnected concepts that can mutually interact and affect one another in multiple ways. The correlation between these concepts is contingent on the specific context and objective of the communication or interaction and may entail complementarity, interdependence, trade-offs, or moderation.

Table 1. The correlation between social presence and virtual presence

	Social presence Virtual presence
Complementarity	Social presence and virtual presence can complement each other to create a more appealing and personalized shopping experience. For instance, social presence can enhance the sense of community and social interaction in virtual shopping environments, while virtual presence can increase the sense of immersion and presence in social environments.
Interdependence	Social presence and virtual presence can depend on each other to enhance the effectiveness of e-commerce. For example, social presence can facilitate the adoption and usage of virtual technologies, while virtual presence can improve social interaction and communication efficiency in online shopping environments[14].
Trade-offs	Social presence and virtual presence can also compete with each other in e-commerce to gain attention and resources [15]. For instance, overemphasizing virtual presence may undermine social interaction and trust-building with other customers or sales representatives, while overemphasizing social presence may distract attention from the immersive and interactive aspects of virtual shopping.
Moderation	Social presence and virtual presence can also moderate each other's effects on consumer behavior in e-commerce. For example, social presence can alleviate the negative effects of virtual anonymity and reduce social anxiety in online shopping, while virtual presence can increase the perceived authenticity and credibility of social information and influence[16].

In summary, virtual presence refers to the sensory and perceptual aspects of the virtual environment, while social presence refers to the social and communicative aspects. Both virtual presence and social presence are crucial for understanding the impact of the virtual environment on individuals and communities and can interact with each other in complex and dynamic ways. Effective utilization of social and virtual presence in e-commerce can enhance customer engagement, increase sales revenue, and improve customer satisfaction and loyalty.

4.2. The difference between social presence and virtual presence

While there are some similarities between social presence and virtual presence, they differ in terms of their definition, focus, technology, and impact.

Table 2. The difference between social presence and virtual presence

	Social presence	Virtual presence
Definition	Social presence refers to the degree to which people feel that they are interacting with real individuals or a group of people during the communication process, even if the communication is mediated through technology[17].	Virtual presence refers to the degree to which an individual feels immersed and present in a virtual environment, giving the sensation of being physically present in the virtual environment[18].
Focus	Social presence in e-commerce focuses on the social connections and sense of interaction in the shopping experience[19], emphasizing the social and communicative aspects of the virtual environment.	Virtual presence in e-commerce emphasizes the sensory and perceptual aspects of the shopping experience[20], such as realism, interactivity, and engagement.
Technology	Social presence in e-commerce can be facilitated through various communication technologies[21], such as chatbots, social media platforms, or online forums.	Virtual presence in e-commerce requires more advanced immersive technologies[22], such as virtual reality, augmented reality, or 3D environments.

Impact	Social presence can influence people's attitudes and behaviors by increasing social connections, reducing social anxiety and improving communication efficiency[23] . The social presence in e-commerce can enhance the social impact, credibility and engagement of the shopping experience.	Virtual presence can have various impacts on human experiences in terms of cognition, emotion, and behavior[24]. In e-commerce, virtual presence can improve product understanding, increase sense of presence, and facilitate decision-making processes.

In summary, social presence and virtual presence in e-commerce are distinct concepts that serve different purposes in enhancing the shopping experience. Social presence focuses on the social aspect of shopping, while virtual presence emphasizes the immersive aspect. Understanding these differences can help retailers develop more effective e-commerce strategies, enhance the shopping experience, and influence consumer behavior in different ways.

5. The impact of social presence and virtual presence on consumer behavior

Level of engagement: Virtual presence and social presence affect the level of consumer engagement with an e-commerce platform[25]. A positive perception of virtual presence and a strong social presence can increase consumers' willingness to spend time exploring the platform, and they are more likely to engage with the platform, participate in discussions, provide feedback or interact, which increases the likelihood of discovering new products and making additional purchases.

Interactivity: When consumers feel they are in a high-quality virtual environment that is easy to interact with in real time and provides a seamless shopping experience, they are more likely to feel connected to the brand and other consumers, and generate positive comments about the brand[26]. This can lead to increased engagement and positive e-word-of-mouth.

Purchase decisions: Perceptions of virtual and social presence can influence consumers' purchase decisions in a variety of ways[27]. For example, a positive perception of virtual presence can enhance consumers' perceptions of product quality and value, increasing their willingness to pay for the product. Similarly, a strong social presence can enhance consumers' perceptions of the trustworthiness and reliability of a platform, thereby increasing the likelihood that they will make a purchase.

Loyalty: Allowing consumers to freely comment, interact and build social identities in real time through immersive experiences, personalized interactions and seamless user interfaces, as well as creating a sense of community around the brand and facilitating customer interactions, all increase consumer trust and loyalty[28]. Loyalty rewards, such as discounts, exclusive content or sneak peeks at new products, are then offered through live streaming to encourage repeat purchases and referrals from consumers to friends.

6. Conclusion

The investigation of social and virtual presence has undergone an evolutionary process, as observed through literature reviews. This journey has advanced from early studies that focused on basic virtual environments to the emergence of virtual reality, in tandem with technological progress and the emergence of novel research inquiries. Although some scholars have made progress in studying user attitudes and behaviors using the concept of virtual presence, research on virtual presence in the emerging and challenging field of live-streaming e-commerce is still in its infancy and urgently needs to be further investigated. Not only does this expand the theoretical research possibilities of virtual presence, but it also furnishes crisis-prone live-streaming e-commerce brands with valuable insights and strategic guidance.

Research shows that social presence and virtual presence are related concepts in e-commerce, with both similarities and differences. Social presence is important for building a sense of connection and rapport between users, helping to create a sense of community and social interaction, fostering a sense of connection and shared experience. It then builds customer trust and loyalty through user-generated content, social media engagement. And virtual presence is important for creating immersion and engagement with the virtual environment, such as through the use of technologies such as augmented reality, virtual reality and real-time streaming, which can increase product satisfaction and purchase intent. In summary, both concepts can be used to enhance the shopping experience and can provide a more immersive and interactive shopping experience for customers, enhancing consumer behavior such as consumer engagement, interactivity, purchase intent, and loyalty[29].

In summary, the study of social presence and virtual presence in e-commerce can yield valuable insights into consumer preferences, needs and pain points, information that can be used to improve products and services, address many questions or concerns, and ultimately provide consumers with a better shopping experience. By gaining a deeper understanding of live e-commerce consumer behavior, e-commerce brands can improve the customer experience and build stronger, more loyal relationships with their customers[30].

References

[1] Bharathy, A. (2021). Sustainable retailing through omni channel strategy: Dawn of a new era. Prayukti – Journal of Management Applications, 01(01). https://doi.org/10.52814/PJMA.2021.1106.

[2] Wohn, D. Y., Freeman, G., & McLaughlin, C. (2018). Explaining viewers' emotional, instrumental, and financial support provision for live streamers. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 1–13. https://doi.org/10.1145/3173574.3174048.

[3] Hu, M., & Chaudhry, S. S. (2020). Enhancing consumer engagement in e-commerce live streaming via relational bonds. Internet Research. https://doi.org/10.1108/INTR-03-2019-0082.

[4] Taher, G. (2021). E-Commerce: Advantages and Limitations. International Journal of Academic Research in Accounting, Finance and Management Sciences, 11(1), Pages 153-165. https://doi.org/10.6007/IJARAFMS/v11-i1/8987.

[5] Wongsunopparat, S., & Deng, B. (2021). Factors Influencing Purchase Decision of Chinese Consumer under Live Streaming E-Commerce Model. Journal of Small Business and Entrepreneurship Development, 9(2). https://doi.org/10.15640/jsbed.v9n2a1.

[6] Bhat, S. A., Kansana, K., & Khan, J. M. (2016). A REVIEW PAPER ON E-COMMERCE. Asian Journal of Technology & Management Research, 6(1), Pages 16-21.

[7] Huang, Z., & Benyoucef, M. (2013). From e-commerce to social commerce: A close look at design features. Electronic Commerce Research and Applications, 12(4), 246–259. https://doi.org/10.1016/j.elerap.2012.12.003.

[8] Cook, A., & Goette, T. (2015). Mobile Electronic Commerce: What Is It? Who Uses It? And Why Use It? Communications of the IIMA, 6(4). https://doi.org/10.58729/1941-6687.1337.

[9] Jeong, H., Yi, Y., & Kim, D. (2022). An Innovative E-Commerce Platform Incorporating Metaverse to Live Commerce (No. 01). ICIC International society. https://doi.org/10.24507/ijicic.18.01.221.

[10] Short, J., Williams, E., & Christie, B. (1976). The social psychology of telecommunications. Toronto; London; New York: Wiley.

[11] Veijalainen, J., & Pitoura, E. (2000). Challenges_in_Mobile_Electronic_Commerce (1).pdf. Proceedings of IeC. https://www.researchgate.net/publication/2431451.

[12] Biocca, F. (1992). Virtual reality technology: A tutorial. Journal of Communication, 42(4), 23–72. https://doi.org/10.1111/j.1460-2466.1992.tb00811.x.

[13] Haimson, O. L., & Tang, J. C. (2017). What makes live events engaging on Facebook Live, Periscope, and Snapchat. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, 48–60. https://doi.org/10.1145/3025453.3025642.

[14] Gao, W., & Li, X. (2019). Building presence in an online shopping website: The role of website quality. Behaviour & Information Technology, 38(1), 28–41. https://doi.org/10.1080/0144929X.2018.1509127.

[15] Chen, J. V., Nguyen, T., & Jaroenwattananon, J. (2021). What drives user engagement behavior in an enterprise SNS account: The role of the INSTAGRAM features. Journal of Electronic Commerce Research, 22(3). http://www.jecr.org/sites/default/files/2021vol22no3_Paper2.pdf.

[16] Hung, S.-W., Cheng, M.-J., & Chiu, P.-C. (2019). Do antecedents of trust and satisfaction promote consumer loyalty in physical and virtual stores? A multi-channel view. Service Business, 13(1), 1–23. https://doi.org/10.1007/s11628-018-0364-y.

[17] Geng, S., Law, K. M. Y., & Niu, B. (2019). Investigating self-directed learning and technology readiness in blending learning environment. International Journal of Educational Technology in Higher Education, 16(1), 17. https://doi.org/10.1186/s41239-019-0147-0.

[18] Mollen, A., & Wilson, H. (2010). Engagement, telepresence and interactivity in online consumer experience: Reconciling scholastic and managerial perspectives. Journal of Business Research, 63(9–10), 919–925. https://doi.org/10.1016/j.jbusres.2009.05.014.

[19] Shen, J. (2012). Social comparison, social presence and enjoyment in the reception of social shopping websites. Journal of Electronic Commerce Research, 13(3), 198–221.

[20] Kapusy, K., & Logo, E. (2017). Values derived from virtual reality shopping experience among generation Z. 2017 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), 000237–000242. https://doi.org/10.1109/CogInfoCom.2017.8268249.

[21] Al-Adwan, A. S., & Kokash, H. (2019). The Driving Forces of Facebook Social Commerce. Journal of Theoretical and Applied Electronic Commerce Research, 14(2), 0–0. https://doi.org/10.4067/S0718-18762019000200103.

[22] C Cummings, J. J., & Bailenson, J. N. (2016). How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence. Media Psychology, 19(2), 272–309. https://doi.org/10.1080/15213269.2015.1015740.

[23] Drouin, M., McDaniel, B. T., Pater, J., & Toscos, T. (2020). How Parents and Their Children Used Social Media and Technology at the Beginning of the COVID-19 Pandemic and Associations with Anxiety. Cyberpsychology, Behavior, and Social Networking, 23(11), 727–736. https://doi.org/10.1089/cyber.2020.0284.

[24] Yung, R., Khoo-Lattimore, C., & Potter, L. E. (2021). Virtual reality and tourism marketing: Conceptualizing a framework on presence, emotion, and intention. Current Issues in Tourism, 24(11), 1505–1525. https://doi.org/10.1080/13683500.2020.1820454.

[25] Algharabat, R., Rana, N. P., Dwivedi, Y. K., Alalwan, A. A., & Qasem, Z. (2018). The effect of telepresence, social presence and involvement on consumer brand engagement: An empirical study of non-profit organizations. Journal of Retailing and Consumer Services, 40, 139–149. https://doi.org/10.1016/j.jretconser.2017.09.011.

[26] Sjöblom, M., Törhönen, M., Hamari, J., & Macey, J. (2019). The ingredients of Twitch streaming: Affordances of game streams. Computers in Human Behavior, 92, 20–28. https://doi.org/10.1016/j.chb.2018.10.012.

[27] Wang, H., Ding, J., Akram, U., Yue, X., & Chen, Y. (2021). An Empirical Study on the Impact of E-Commerce Live Features on Consumers' Purchase Intention: From the Perspective of Flow Experience and Social Presence. Information, 12(8), 324. https://doi.org/10.3390/info12080324.

[28] Prahiawan, W., Fahlevi, M., Juliana, J., Purba, J. T., & Tarigan, S. A. (2021). The role of e-satisfaction, e-word of mouth and e-trust on repurchase intention of online shop. International Journal of Data and Network Science, 5(4), 593–600. https://doi.org/10.5267/j.ijdns.2021.8.008.

[29] Dubosc, C., Gorisse, G., Christmann, O., Fleury, S., Poinsot, K., & Richir, S. (2021). Impact of Avatar Anthropomorphism and Task Type on Social Presence in Immersive Collaborative Virtual Environments. 2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), 438–439. https://doi.org/10.1109/VRW52623.2021.00101.

[30] Shin, D.-H., & Shin, Y.-J. (2011). Consumers' Trust in Virtual Mall Shopping: The Role of Social Presence and Perceived Security. International Journal of Human-Computer Interaction, 27(5), 450–475. https://doi.org/10.1080/10447318.2011.552060.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230170

The History and Current Situation of China's Undergraduate Major Evaluation and the Prospect of Intelligent Education Evaluation Based on Big Data Technology

Shiqian GU[a] , Wei LI[b,1] and Xiaoli LIU[c,2]
[a] *Education Research Institute, Xiamen University, Xiamen, China. 361005*
[b] *Undergraduate School, Jinan University, Guangzhou, P. R. China. 510632*
[c] *College of Information Science and Technology, Jinan University, Guangzhou, China. 510632*

Abstract. The accreditation and evaluation of undergraduate majors is one of the most important elements in the monitoring of China's higher education quality. with the combination of quantitative and qualitative research methods, this paper illustrates the stages of development, problems of undergraduate majors operation in China, and proposes an evaluation framework for undergraduate majors from the fourth paradigm perspective. The development of undergraduate major evaluation in China are divided into four stages: the budding period (1985-1998), the rising period (1998-2009), and the booming period (2010-2022). The forms of evaluation are divided into major accreditation, major ranking and major assessment. Major evaluation mode includes independent evaluation mode, comprehensive evaluation mode and appraisal mode. Continuous collection of sample data, customized indicators, multiple fusion calculation analysis, visual feedback are the typical features of big- data-based intelligent education evaluation.

Keywords. Undergraduate Major Evaluation, History and Current Situation, Big Data

1. Introduction

The accreditation and evaluation of undergraduate majors is one of the most important elements in quality control of higher education in China. Why is it necessary to carry out undergraduate major assessment? Undergraduate major is the basic element in universities and the basic unit for universities to realize their functions. The pursuit of high-quality major education is a strong motivation in the development of major evaluation. With the deepening of reform in education sector and the loosening on major setting as well as approval restrictions, universities thus are more active in setting majors independently. Therefore, it is critical to ensure high level of major setting and

[1] Author a and b contributed equally to this work.

[2] Corresponding Author, LIU Xiaoli, College of Information Science and Technology, Jinan University, Guangzhou, P. R. China; E-mail: txlliu@jnu.edu.cn.

This research is supported by Guangdong Education Science Planning Project (2022GXJK141), National Natural Science Foundation of China (Grant No.82074331).

management as well as high quality of education with professional standards and evaluation system. Majors of higher education link universities and society. However, the incompatibility between the supply of majors and the demand of economic and social development still prevails, and the contradiction between structural unemployment of graduates and enterprises' failure to find suitable talents is still to be solved. Therefore, it is necessary to form a mechanism for universities to develop mature schemes of self-building, self-development and self-improvement in major management through goal-oriented assessment. The quality of talents cultivation and their adaptation to social demands also need to be further improved. This paper systematically analyzes the theoretical and practical development and problems of undergraduate major evaluation in China, and proposes the theoretical prospects and practical strategies for the development of major evaluation in the future.

2. Study on the development stages of major evaluation in China

Adopting both quantitative and qualitative research methods, this paper takes references from the database of China Knowledge Network Literature (CNKI) and visual analysis on China's major evaluation research literature by the knowledge mapping analysis software Cite Space[3] and the CNKI's measurement and visualization function. Through a general study of the mapping of major evaluation research fields, the research results of major evaluation in China are unfolded in a holistic and multi-faceted manner in this paper, representing the theoretical development of major evaluation as a whole.

Johannes Kepler discovered the laws of planetary motion based on Tycho Brahe's systematic observations on celestial motions. Likewise, Li Jie also put up that the focus and paradigm of scientific research changes over time, sometimes slowly and sometimes dramatically. And the history of scientific development can be tracked from published literature [1]. Being inspired by these examples, the author applies to the date from literature, the number of published articles, keywords, co-citations, emergent terms, authors and institutions to outlining the panorama of major evaluation in depth and details.

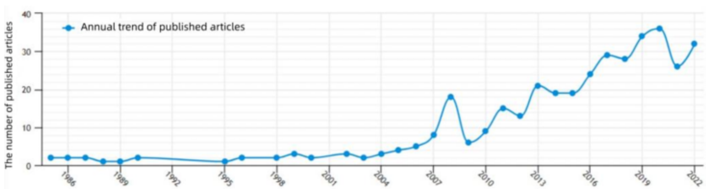


Figure 1. The trend of published journals about major evaluation and major assessment from 1985-2022

As what can be seen from Figure 1, research on major evaluation in China began in 1985. For more than 20 years after the first journal article on major evaluation was published, the number of relevant published journal articles has been remained at the level of about 1-5 per year, drawing few attentions from scholars who focus on the theoretical research of higher education. Overall, the development of research related to major evaluation can be divided into three periods: budding, rising and booming.

[3] Cite Space (Citation Space) is a software for visualizing and analyzing trends and patterns in scientific literature. It is designed as a tool for dig out the clustering and distribution of knowledge in cited literature through analysis on citation from internet.

2.1. The first period: budding (1985-1998)

The issuing of *The Decision of the Central Committee of the Communist Party of China on the Reform in Education System* in May of 1985 marked China's education system reform beginning. With stronger attention to the quality of undergraduate education teaching and talent cultivation from central government, China launched education evaluation projects from nation, province and university level, when the exploration on major evaluation theory sprouted. However, the number of articles published at that time was still at a low level of 1-2 articles each year. At this period, related research topics mainly focus on the preliminary exploration of evaluation theory and practice, most of which are empirical articles.[2-4]

2.2. The second period: rising (1998-2009)

According to The Law on Higher Education promulgated at the fourth meeting of the Standing Committee of the 13th National People's Congress in August of 1998, the level and quality of education in universities and colleges have to be supervised and evaluated by educational administrative departments, offering legal support for educational evaluation's importance and normativity. What came after is the increasing of the theoretical research on major evaluation. During the same time, the number of published journals increased as a whole, though with occasional dropping in some years. 2008 saw the highest point of 18 journals published while a downturn occurred in 2009. At this period, the research and practice of major evaluation began to learn international experience[5-6], attach importance to evaluation methods[7-8] and the guidance of relevant theories.[9]

2.3. The third period: booming (2010-2022)

Generally speaking, the number of articles published during this period shows a wavy but upward trend. Derived by the issuing of *The Guidance on Accelerating the Implement of "Double First-class" Initiative in Higher Education Institutions* jointly formulated by the Ministry of Education, the Ministry of Finance and the National Development and Reform Commission in August of 2018 and The General Plan for Deepening the Education Evaluation Reform in the New Era issued by the Central Committee of the Communist Party of China and the State Council in October of 2020, the number of published journals peaked with 36 in 2020 alone. By 2022, the published journals on major evaluation in higher education disciplines has accumulated to 351. At this period, relevant research is constantly seeking how to improve major evaluation. [10]

On the University's Competitiveness and Evaluation of University's Specialty by Zhang Xiaodan from Wuhan University is the first dissertation on major evaluation. This dissertation proved the scientificity and practicality of major evaluation index system from the perspective of empirical evidence. The purpose of the research is to systematically elaborate major evaluation system and use its index system to promote the competitiveness of universities, with the establishment of various index systems and the study on evaluation system of major classification in universities at its core. [11] As can be seen from Figure 2, from 2004 to now, the dissertations with the title of major evaluation have reached 49. The highest number of 8 was in 2010 and the number has been maintained at about 4 in the past two years.

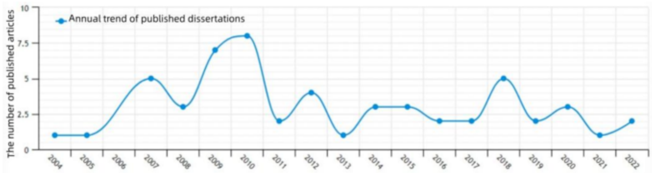


Figure 2. The trend of dissertations published with the title of major evaluation from 2004-2022

3. Current status of major evaluation practice

3.1. Forms of major evaluation

In recent years, the rapid and widespread development of higher education has drawn worldwide attention on higher education assessment. In terms of the form of major evaluation, there are currently three main types (As shown in Figure 3).

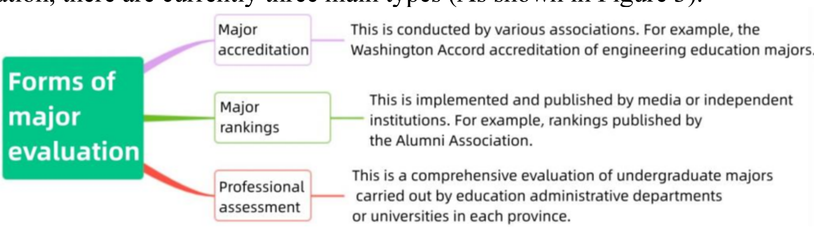


Figure 3. Forms of major evaluation in China

One is major accreditation, which is conducted by various associations. For example, the Washington Accord accreditation of engineering education majors.

The second is major ranking, which is implemented and published by media or independent institutions. For example, rankings published by the Alumni Association (As shown in Table 1).

Table 1. Alumni Association's 2017 ranking of undergraduate majors in Chinese universities - Guangdong Province

Ranking	University	Type	National Ranking	8 Star	7 Star	6 Star	5 Star	4 Star	3 Star	2 Star	1 Star	Sum
1	Sun Yat-sen University	Multiversity	17	0	5	10	28	59	16	0	0	118
2	South China University of Technology	Science and engineering university	45	0	1	4	11	41	33	7	0	97
3	South China Normal University	Normal university	91	0	0	3	5	13	34	20	0	81

The third is professional assessment, which is a comprehensive evaluation of undergraduate majors carried out by education administrative departments or universities in each province. As of 2022, Guangdong, Liaoning, Shanxi, Hubei and other administrative departments in education sector have carried out provincial undergraduate major evaluations.

3.2. Mode of major evaluation

When it comes to evaluation modes, they can be classified into three modes: independent evaluation mode, comprehensive evaluation mode and appraisal mode (As shown in Figure 4).

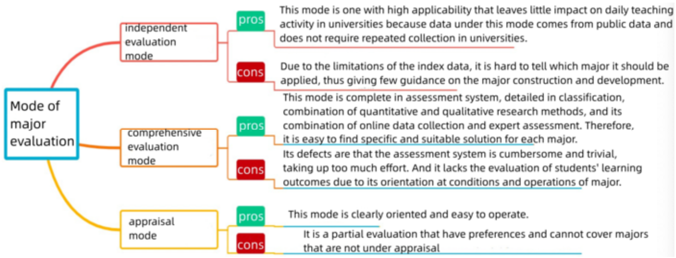


Figure 4. Analysis of the classification of major evaluation modes in China

The first is independent evaluation mode. Being commissioned by the Department of Higher Education of the Ministry of Education of the People’s Republic of China, the project Analysis on the Evaluation and Star Distribution of the Disciplines and Majors of Project 985 Universities (National Key Universities) dominated by Professor Qiu Junping is an independent evaluation mode. This project studied the quality of the construction of majors in 985 Universities from 2012 to 2013. The raw data in this evaluation mainly come from official data documents (compilation, yearbook, report, etc.).

The second is comprehensive assessment mode. Taking provincial undergraduate program evaluation in Liaoning Province for example, there are six ground rules in their evaluation. The first is orientation. The practice of comprehensive evaluation of majors further promotes universities to take undergraduate teaching as fundamental work, so that dean’s and faculty’s focus, resource allocation and funding arrangement are all fixed on teaching. Such evaluation can also continuously strengthen major construction and reform, improve the level of major construction and talent quality, and advance majors to better serve for economic and social development in a faster step. The second is scientification. The design of the comprehensive major evaluation index system, the choice of evaluation methods and evaluation practice should follow the law of education and the law of professionals development, fully consider the inherent characteristics of major construction and talent training and effectively promote the comprehensive, coordinated and sustainable development of major. The third is objectivity. Through comprehensive evaluation of majors, representative and repeatable indicators are selected among indicators reflecting the state of major management. And a scientific and reasonable indicator system and evaluation function are designed to objectively reflect the real state of reform and construction of majors. The fourth is simplicity. The comprehensive major evaluation resorts to modern information technology means, which helps to collect, summarize and analyze data through internet, to simplify evaluation process and improve efficiency, instead of doing research in universities by experts themselves. The fifth is offering specific guidance for different majors. Comprehensive major evaluation is to evaluate the same majors offered by different schools. On the basis of not affecting normality of major management, the designing of index system and the application of evaluation results is used for providing specific guidance for each major to help them develop with their own characteristics. The sixth is the mixed quantitative and qualitative research methods. The comprehensive evaluation of majors adopts the quantitative analysis of data as the main focus and the qualitative judgment from experts as the supplement. While the former focuses on the objective evaluation of the current state, the latter highlights the subjective evaluation on potentials. The second is comprehensive assessment mode. Taking provincial undergraduate major evaluation began in Liaoning Province for example, there are six ground rules in their evaluation.

The first is orientation. The second rule is scientification. The third is objectivity. The fourth is simplicity. The fifth is offering specific guidance for different majors. The sixth is combining quantitative and qualitative research methods.

The third is appraisal mode. In 2016, the General Office of Shanxi Provincial Party Committee and the General Office of Shaanxi Provincial People's Government decided to launch the First-class Majors Initiative. Project management is used in the building of first-class majors. All undergraduate colleges and universities in the province are required to apply their majors to step into first-class major. The Shanxi Provincial Department of Education entrusted the Western China Higher Education Assessment Centre with the responsibility of accepting materials and auditing projects for undergraduates, and the Shanxi Higher Education Data Centre with the responsibility of data collection and technical support. Shanxi Vocational and Technical Education Society was entrusted by the Provincial Department of Education with the responsibility of accepting materials and auditing projects for higher education.

Based on what is analyzed above, it can be concluded that independent evaluation is one with high applicability that leaves little impact on daily teaching activity in universities because data under this mode comes from public data and does not require repeated collection in universities. However, at the same time, due to the limitations of the index data, it is hard to tell which major it should be applied, thus giving few guidance on the major construction and development. The comprehensive evaluation mode stands out with its completeness in assessment system, detailed in classification, combination of quantitative and qualitative research methods, and its combination of online data collection and expert assessment. Therefore, it is easy to find specific and suitable solution for each major. However, its defects are that the assessment system is cumbersome and trivial, taking up too much effort. And it lacks the evaluation of students' learning outcomes due to its orientation at conditions and operations of major. The appraisal mode, on the other hand, is clearly oriented and easy to operate, but it is an evaluation that only focus on majors that are under appraisal. So, it is a partial evaluation that have preferences and cannot cover majors that are not under appraisal.

The preceding analysis on these pros and cons of the three modes has inspired the author to make further step on the design of major evaluation. From the perspective of promoting evaluation discernment, major evaluation can put more emphasis on the orientation of indicators, strong data support, appropriate assessment scales and high universality in majors that are being evaluated. In terms of convenience, the data will be judged mathematically and processed with big data technology. This procedure ought to be responsible by an outstanding team, rather than be finished annually with qualitative assessment by a large number of experts, resulting in low efficiency due to different standards. In terms of orientation, this type of evaluation will cover as many majors as possible, giving the same attention on both students' learning outcomes and keeping characteristics of each major.

4. Analysis of the problems of major evaluation

In October 2020, the Central Committee of the Communist Party of China (CPC) and the State Council issued the General Plan for Deepening the Reform of Education Evaluation in a New Era, which calls for improving the institutional mechanism of moral education, reversing the unscientific orientation of education evaluation, and resolutely overcoming chronic problems of centering at higher scores better admissions, more paper, higher career status. In practice, major evaluation has always been one of the difficulties in evaluating higher education and a headache for education administrators and evaluators. [12] It is of strategic importance to deepen the reform of education evaluation in the new era by innovating the concept of undergraduate major evaluation and constructing a new system and mode of major evaluation.

The big-data analysis on Chinese discipline major research based on the study of China's core journals reveals that China's major evaluation research on disciplines outweigh the research on majors. More vividly, Figure 5 illustrates the big data cloud map and knowledge map of specific research keywords. In-depth analysis have made it clear that China's higher education evaluation has gone through four stages: beginning, institutionalization, advancing and deepening. Since the reform of the education system in 1985, China has put more stress on the quality of undergraduate education teaching and major talent cultivation, and launched relevant evaluation projects at three levels: national, provincial and university. A system of higher education evaluation with Chinese characteristics have basically formed after 40 years' exploration and practice. In the new era, the Double First-class Initiative has been officially implemented, but major evaluation has not been adapted to it, and there are still problems such as putting more emphasis on macro evaluation than micro evaluation, obvious utilitarian orientation, more quantitative and result-oriented indicators but less qualitative and process-oriented indicators, lack of big data support for evaluation, convergence of evaluation indicators and lack of discipline and major characteristics. [13]

Figure 5. Big data cloud map and knowledge map of keywords for major evaluation research in Chinese disciplines

In summary, theoretical research on major evaluation in China began in 1985 with the reform of the national education system, and has gone through three stages: budding, rising and booming. Throughout the 37 years of theoretical development, the unswerving efforts of Chinese scholars in exploring major evaluation theory have also promoted the development of Chinese educational evaluation theory, provided theoretical support and paradigm for the practice of major evaluation in the field of higher education, and made theoretical contributions to the modernization of Chinese-style education. However, the number of publications and of representative research results is few, which needs to be further explored.

5. Prospects: A Framework for Undergraduate major Evaluation in the Fourth Paradigm Perspective - Smart Education Evaluation Based on Big Data Technology

The surging of the big data technology has unprecedentedly advanced things to be quantified and statistically analyzed, giving rise to the new progress on data-intensive knowledge and driving evaluation paradigm's evolution. According to Turing Award winner Jim Gray, scientific research can be divided into four paradigms: the Experimental Science paradigm for describing natural phenomena (Empirical Science), the Theoretical Science paradigm using modeling method and inductive method (Theoretical Science), the Computational Science paradigm for computer simulations of complex phenomena (Computational Science), and the data-intensive science paradigm (eScience), also known as the fourth paradigm, which combines theory, experiment and computational simulation. The fourth paradigm includes the usage of diverse tools for the continuous collection of scientific data, building system for managing the entire lifecycle of data, and the designing of tools and methods for data analysis and visualization customized for scientific research questions. [14] (As shown in Figure 6). There is no doubt that the advent of big data technology provides new methods and new perspectives for our cognition and scientific research. If this technology is applied in undergraduate major evaluation, a new paradigm of data-intensive evaluation will be invented and help undergraduate major evaluation to be conducted in an intelligent way. In other words, data mining and learning analysis based on full-sample, full-process and panoramic meta-education data will enable the evaluation of undergraduate majors to become intelligent.

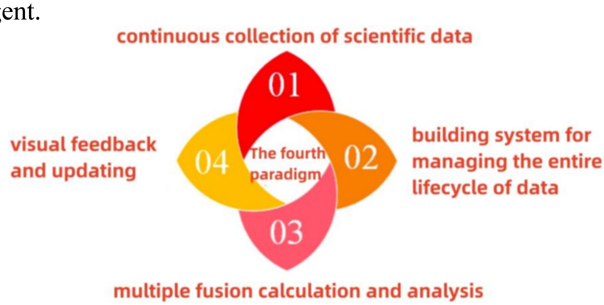


Figure 6. Main features of the fourth paradigm

To be specific, on account of applying big data technology to education research, education evaluation methods would see revolution, evolving into a data-intensive dynamic evaluation system that combines uninterrupted collection of sample data, personalized indicators, multiple fusion calculation and analysis, as well as visual feedback and updating together.

In the future, major evaluation will use the Internet of Things perception technology, video recording technology, image recognition technology and platform acquisition technology to continuously collect the multi-source, heterogeneous, multi-modal and incoherent semantic big data generated in real time in the process of education. major evaluation activities are no longer limited by presets and are not limited to the investigation of causality. Instead, specific algorithms are used to analyze "big data", so as to conclude the correlation and regularity behind the educational data. In the future major evaluation, big data technology can mine valuable information from the big data of education generated in real time through the real-time dynamic monitoring of educational activities, and present the educational evaluation results in intuitive graphics and image information through visual tools, and personalized feedback the evaluation results to the evaluators. Big data technology has continuously improved the professional level of educational evaluation.

6. Summary

Focusing on the evaluation of undergraduate majors in China, this paper clears out three phases of undergraduate major development in China: budding, rising and booming, by employing quantitative and qualitative research methods. Besides, this paper also classifies the forms of major evaluation in China into major accreditation, major ranking and major assessment, and the major evaluation modes into independent evaluation mode, comprehensive evaluation mode and appraisal mode. Furthermore, the fourth paradigm of evaluation framework of undergraduate major is proposed in this paper, with a view to make up for the shortcomings of previous research methods and provide references for research and practice related to the innovation of undergraduate program evaluation system.

References

[1] Li J, Chen C M, CiteSpace:Text Mining and Visualization in Scientific Literature (Second Edition), *Capital Economic and Trade University Press*, pp.02, 2017.
[2] Xie B Z, On the practice of discipline and specialty evaluation in colleges and universities, *Research in Educational Development*, vol.02, pp. 43-47, 1990. (in Chinese)
[3] Gan F, Chen L J, A rustic opinion of the evaluation of university running level, *Research in Educational Development*, vol.01, pp. 57-63, 1986. (in Chinese)
[4] Cao S H, Chen S H, An attempt has been made by Tongji University to assess majors, *Research in Educational Development*, vol.04, pp. 100-103, 1985. (in Chinese)
[5] Mao X Y, An analysis of the application of professional development schools in teacher education - An assessment of the Teaching Master's Program at Trinity University, *Teacher Education Research*, vol.01, pp. 73-78, 1990. (in Chinese)
[6] Yang X Q, Wang G X, Tian B X, The recognition system for professional education in US higher engineering education and its implication, *Research in Higher Education of Engineering*, vol.01, pp. 76-78, 2005. (in Chinese)
[7] Zhao T T, Application of comprehensive value assessment method in evaluating colleges and universities' professional programs, *Journal of Technology Economics*, vol.03, pp. 61-63, 1999. (in Chinese)
[8] Dong X H, Professional Accreditation: An Approach to Quality Assurance in Higher Education, *Fudan Education Forum*, vol.06, pp. 33-38, 2008. (in Chinese)
[9] Ni X M, Professional evaluation: Quality assurance for the social dimension, *Jiangsu Higher Education*, vol.04, pp. 29-31+77, 2004. (in Chinese)
[10] Yu T, Liu S X, From External Assessment to Self-improvement: The Reform of EC2000 and lts lmplications, *Research in Higher Education of Engineering*, vol.06, pp. 28-34, 2014. (in Chinese)
[11] Zhang X D. On the University's competitiveness and Evaluation of University's specialty. *Wuhan University*, pp.03, 2004. (in Chinese)
[12] Li M L, Wang Z J. The Evaluation of First-class Major Construction in the New Era Should be Effectiveness-based. *Jiangsu Higher Educatio*n, vol.09, pp. 20-23, 2020. (in Chinese)
[13] YANG P P, Wang X. Research on the innovation of discipline evaluation in China under the background of 'Double First-Class' construction. *Journal of Higher Education Management*, vol.06, pp. 65-73, 2018. (in Chinese)
[14] Hey, T., Tansley S. & Tolle, K. The Fourth Paradigm: Data-intensive Scientific Discovery. Pan Jiaofeng, Zhang Xiaolin.etc . *China Science Publishing & Media Ltd.(CSPM)*, pp.01,2012.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230171

A Probabilistic Guide for Domestic Stocks Delisting Risk from the Nature of Bayesian Matrix

Runnan Chen and Lei Zhang[1]
School of Economics and Management, Beijing Jiaotong University, Beijing 10044, China

Abstract. Finding an optimum way to identify stocks with less delisting risk is critical for every investor in the stock market. However, this procedure is often done based on personal experience, which doesn't fully utilize the historical delisting records. This convention of selecting stocks might result in a greater loss since it merely involves subjective judgment, especially for individual investors. Our research proposes a probabilistic approach for identifying the delisting risk associated with different industry sectors, given the P/B ratio level distribution. And this research offers a customized guide for individual investors to better choose the safer investment options related to the stocks' industry sectors. The completion of our conditional probability matrix is operated under the high-rank assumption, together with the features of Bayesian matrices. The experimental results for our domestic delisting stocks supports the validity and usefulness of our method.

Keywords. Matrix completion, delisting risk, Bayesian matrix, domestic stock data

1. Introduction

The delisting of stocks can bring detrimental impact on their current stockholders: decreased liquidity, less protections toward firms' frauds and increased risk of investing [1, 2]. However, with the little access to investigate the firm's internal operating status, it's hard for stockholders to learn their chances of delisting. Therefore, finding a proper criterion to evaluate the possibilities of delisting is crucial. And the price-to-book ratio (P/B ratio) has been a long-standing important indicator using in industry-wide stock risk evaluation [3, 4, 5]. Noticeably, the interpretation of this numeric ratio is highly dependent on the industry sector of the stock. Hence, considering the availability and usefulness of the historical information, a P/B ratio-based delisting evaluation offers practical methods for stockholders to select stock in different industrial sectors. As for individual investors, this selection process primarily relies on their personal analysis of the historical data of various indicators. The delisting risk evaluation can be highly biased among these subjective opinions. Focusing on China's domestic stock market, there's still no probabilistic guide about the delisting risk while delisting firms' information is avail. With a probabilistic guide regarding different delisting risk and related P/B ratio among different sectors, investors can make a more cautious choice.

[1] Corresponding Author, Zhang Lei, School of Economics and Managements, Beijing Jiaotong University, Beijing, 10044, China; E-mail: zhlei@bjtu.edu.cn.

Current research on delisting prediction majorly utilizes regression model and machine learning (ML) methods: Hwang et al. use a set of both nonfinancial and financial indicators to build logit regression model, and emphasizing an equal attention for both kinds of indicators [6]; Zhou applies Adaboost method to perform simple classification task on firms' delisting, which achieves a better accuracy than other existing classification methods [7]. Meanwhile, current investment-related research on matrix completion focuses on low-rank assumption or periodical behavior: Agrawal reviews existing methods for matrix completion, and shows the Singular Value Decomposition using the Order Method is the best to complete missing stock indices [8]; Athey et al. develop a series of matrix completion estimators to allow missing values have time-series pattern [9]. Previous techniques have been widely used in stock-related information prediction and completion. However, the previous research on prediction of the delisting risk cannot provide a stochastic reference for stock selection. Under the specific scenario, the P/B ratio and industrial sector distribution of the delisting stock form a high-rank matrix, which violates the low-rank assumption. Thus, our research proposes a novel approach to extract probabilistic information regarding the P/B ratio and industry-dependent delisting risk, aiming at better supporting domestic stockholders investing decisions.

2. The construction of the industry-PB ratio Bayesian matrix

From the data source, we can obtain the distribution of industry sectors of all delisted stocks: manufacturing, IT, real estate, and the other 13 sectors. For any individual stockholder, he/she might choose a portfolio of stocks, which is consisted of diverse industry sectors. It is worth noticing that, even the P/B ratio can be a generalized metric to evaluate the quality of stock, we can hardly interpret it directly by merely referring to its numeric value. So, we develop three predefined thresholds for benchmarking this ratio based on industry-wide convention [10]: low (P/B ratio less than 1), medium (P/B ratio falls between 1 and 3), high (P/B ratio falls between 3 and 10), extreme (P/B ratio greater than 10). Set the thresholds as $V = (v_1, v_2 \ldots v_x)$, industry sectors as $H = (h_1, h_2 \ldots h_y)$. And vector H should be initially given by the distribution of historic delisted data specified by different stock market.

The distribution of the delisted industry sectors is associated with the stocks' performance in the P/B ratio. Therefore, we can express their relationship in the form of conditional probabilities: $P(H|V)$ is the probability of the delisted stock belongs to the ith industry sector given its P/B ratio threshold. And $P(V|H)$ is the probability of the stock has jth P/B ratio threshold given its industry sector. We can write the two conditional probabilities information into a Bayesian matrix, M:

$$M = \begin{pmatrix} 0 & P(V|H) \\ p(H|V) & 0 \end{pmatrix}, V = \begin{pmatrix} v_1 \\ \cdot \\ \cdot \\ \cdot \\ v_x \end{pmatrix}, H = \begin{pmatrix} h_1 \\ \cdot \\ \cdot \\ \cdot \\ h_y \end{pmatrix} \tag{1}$$

We partition this matrix into four areas: the two areas that contain non-zero values are lower-left and upper-right: (1) the probabilities that the stock comes from this industry sector given its P/B ratio level. (2) the probabilities that the stock has this P/B ratio level given its industry sector. The two areas that are all zero-valued are upper-left and lower-right: since we set (3) there is no linkage between different stocks' industry sector and (4) there is no correlation between their P/B ratio level, the conditional probabilities will always be zero. Therefore, to solve this objective matrix M, we need to complement the lower-left and upper-right parts.

3. Complementing Bayesian matrix with high-rank settings

Because every element in our industry-PB matrix, M, represents for a certain probability, so they must fall in between the range from 0 to 1. Also, based on our previous settings that each column displays the all circumstances of conditional probabilities, so the column-wise sums of the industry-PB matrix will all equal to 1. After deriving these properties, the industry-PB matrix can be considered as a Markov matrix. And this industry-PB matrix is inherently possesses an eigenvector with eigenvalue equals to 1 because of the Markov matrix's feature. Now we need to find this eigenvector. Consider Equation (2) below:

$$\begin{pmatrix} 0 & P(V|H) \\ p(H|V) & 0 \end{pmatrix} \begin{pmatrix} P(V) \\ p(H) \end{pmatrix} = \begin{pmatrix} P(V) \\ p(H) \end{pmatrix} \tag{2}$$

This equation proves that, vector $r = \begin{pmatrix} p(V) \\ p(H) \end{pmatrix}$ is the corresponding eigenvector with the eigenvalue 1 of our industry-PB matrix, alternatively we can call it the principal eigenvector. Placing it in the scenario of delisted stock analysis, it consists of two components: the probability distribution of the P/B ratio levels and the probability distribution of the industry sectors among all delisted stocks. After getting this eigenvector, we need to reversely infer the original matrix elements. Consistent with the basic assumption that, there's great uncertainty and fluctuation around the coverage of every industry sector and every P/B ratio level, we discuss the finding process under high rank Markov matrix condition. So, our industry-PB matrix has a steady state after multiplying itself n times:

$$\lim_{n \to \infty} M^n = R = (r, r, \cdots, r), r = \begin{pmatrix} r_1 \\ \cdot \\ \cdot \\ \cdot \\ r_{x+y} \end{pmatrix} \tag{3}$$

With

$$MR = RM = R \tag{4}$$

Integrating these equations, we can express the eigenvalue λ by using matrix M:

$$\frac{r^T MRMr}{r^T r} = \frac{r^T Rr}{r^T r} = \lambda \tag{5}$$

So, the problem becomes an optimization problem. i.e., find the matrix M that maximizes the formula (5):

$$Maxmize \ \frac{r^T MRMr}{r^T r}$$

Given constraints (m_{ij} is the element of M at ith row, jth column; r_i is the ith element of vector r):

$$0 \leq m_{ij} \leq 1$$
$$\sum_{i=1}^{x+y} m_{ij} = 1$$
$$m_{ij}^3 r_j = m_{ji}^2 r_i$$

Eventually, we can solve a matrix M, which subordinates to all constraints above and conforms with the previous eigenvector r. Therefore, with the information drawing from a posterior probabilistic vector r, individual investors can be more well-informed about the delisting risk while knowing stocks' P/B ratio in particular industry.

4. Experiment

Using the delisted stocks data available on Shanghai Stock Exchange and Shenzhen Stock Exchange websites [11, 12], we derive the distribution of delisted stocks' industry sectors: we denote manufacturing as h1, IT as h2…and the rest of industry sectors following the same name convention, which showing with their corresponding percentages in the Figure 1. Also, among the total 124 delisted stock records, we can categorize their P/B ratio into four thresholds: v_1 (low): 29.8%, v_2 (medium): 27.4%, v_3 (high): 29%, v_4 (extreme): 13.7%. From this information, we can construct the eigenvector of our Bayesian matrix: (0.298387097, 0.274193548, 0.290322581, 0.137096774, 0.532258065, 0.088709677, 0.064516129, 0.056451613, 0.048387097, 0.040322581, 0.032258065, 0.032258065, 0.024193548, 0.024193548, 0.016129032, 0.008064516, 0.008064516, 0.008064516, 0.008064516, 0.008064516)^T.

Figure 1. Distribution of stocks' industry sector.

Utilizing the nonlinear optimization model developed in Section3, we obtain the complements of the upper right and lower left entries of the Bayesian matrix from our eigenvectors. The detailed information of the matrix entries and the complement results are displayed in Figure 2 and Figure 3. We use the value in row 5, column 1, to illustrate how an investor can utilize the results: it informs the investor that a delisted stock with a low P/B ratio level has a 58.79% probability that it comes from the manufacturing industry, h1. Compared with other entries located in column 1, this probability is significantly high. Now, the investor might become more cautious about selecting stocks from the manufacturing industry with a low P/B ratio for his/her portfolio.

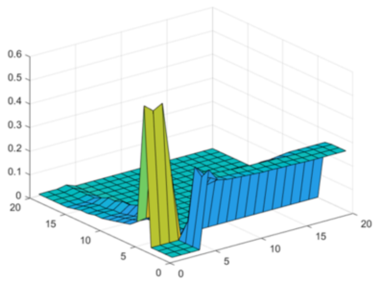


Figure 2. Values of the matrix entries.

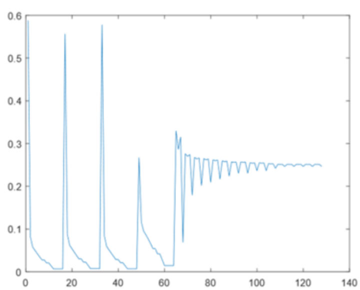


Figure 3. Complement results.

5. Conclusion

This paper provides a basic algorithm to extract the probabilistic relation between P/B ratio levels and delisting risk given the stock's industry sector using properties of Markov matrices by forming an eigenvector containing the distribution of industry sectors and the distribution of P/B ratio levels. It can serve as a probabilistic guide for individual investors to choose safer stocks regarding the delisting risk. The experiment implementation validates the viability of the approach we offered in this paper. The future research can be refined in different aspects: (1) the selection of more comprehensive delisting risk metrics; (2) the incorporation of dynamic risk prediction using time-series information.

Acknowledge

The work was supported by the National Natural Science Foundation of China (No.72271017).

References

[1] Macey J, O'Hara M, Pompilio D. Down and out in the stock market: the law and economics of the delisting process[J]. The Journal of Law and Economics, 2008, 51(4): 683-713.
[2] Loureiro G, Silva S. The impact of cross-delisting from the US On firms' financial constraints[J]. Journal of Business Research, 2020, 108: 132-146.
[3] Cheng C S A, McNamara R. The valuation accuracy of the price-earnings and price-book benchmark valuation methods[J]. Review of quantitative finance and accounting, 2000, 15: 349-370.
[4] Shittu I, Ahmad A C, Ishak Z. Price to book value, price to sales multiples and stock price; evidence from Nigerian listed firms[J]. Journal of Advanced Research in Business and Management Studies, 2016, 3(1): 85-93.
[5] Leonardo M P, Eduardo C M C, María T H A, et al. Formalization of a new stock trend prediction methodology based on the sector price book value for the Colombian market[J]. Heliyon, 2022: e09210.
[6] Hwang I T, Kang S M, Jin S J. A delisting prediction model based on nonfinancial information[J]. Asia-Pacific Journal of Accounting & Economics, 2014, 21(3): 328-347.
[7] Zhou L. Predicting the removal of special treatment or delisting risk warning for listed company in China with Adaboost[J]. Procedia Computer Science, 2013, 17: 633-640.
[8] Agrawal P. Predicting Stock Market Indices Using LRMC[J]. 2021.
[9] Athey S, Bayati M, Doudchenko N, et al. Matrix completion methods for causal panel data models[J]. Journal of the American Statistical Association, 2021, 116(536): 1716-1730.
[10] Pae J, Thornton D B, Welker M. The link between earnings conservatism and the price-to-book ratio[J]. Contemporary Accounting Research, 2005, 22(3): 693-717.
[11] Shanghai Stock Exchange. Suspension[EB/OL]. [2023-02-08]. http://www.sse.com.cn/disclosure/dealinstruc/suspension/.
[12] Shenzhen Stock Exchange. Market Data[EB/OL]. [2023-02-08]. https://www.szse.cn/market/stock/suspend/index.html.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230172

Sharing and Co-Construction of Library Alliance Driven by Big Data and Blockchain

Jing Ling[1] and Lidong Huang
Library of Xiangnan University, Chenzhou, 423000, China

Abstract. Big data technology drives the library alliance and shifts focus from the construction of cooperation mechanism to the construction of technology level. It integrates library big data by building a national shared service platform covering concepts, resources, technology, readers, mechanisms and then fully share the literature resources within the library alliance with the help of blockchains and alliance chains technology to explore the future development and direction of the library.

Keywords. Library, library alliance, big data, blockchain, co-construction, sharing

1. Introduction

Before the advent of computers, libraries were the only sources of literature review and data collection for the researchers. Subsequently, the computer applications enabled easily to compile machine readable catalog (MARC) data in order to retreat library's massive literature information resources. The use of Internet further exploded the network information and the position of libraries as an information center affected significantly. The emergence of the search engines had rapidly replaced libraries as the first choice of the people to obtain information and libraries were gradually marginalized. With the development of big data and blockchain technology, massive data and advanced algorithms enables scholars to search relevant information. There are two laws in this series of changes. First being the changes in the way the people obtain information have changed with the progress of science and technology, that is, "productivity determines superstructure". Second, the speed by which people access information is constantly accelerating, and the access speed has become a key factor in the survival of the fittest. In the face of the society's ever-increasing demand for accessing information, no single library can meet their expectations, but even libraries all over the country or the world would be inadequate. The World Library will be the ultimate ideal form that will emerge in near future.

[1] Corresponding Author, Jing Ling, Library of Xiangnan University, Chenzhou, 423000, China; Email: 125435785@qq.com

2. Current situation of domestic library alliances

Library alliance refers to a library consortium in which several libraries sign a cooperation agreement for the purpose of co-construction and sharing of literature information resources [1]. Geographically, the forms of the library alliances can be divided into national alliances and regional alliances according to the scope covered. It can also be divided into the same system alliances or cross system alliances according to the types of the libraries. The specific forms include national same system alliances, national cross system alliances, regional same system alliances, and regional cross system alliances [2].

China's library alliance is dominated by the university libraries and the public libraries, such as China's Higher Education Document Security System, Hunan University Digital Library, etc. [3]. Regional library alliances led by the public libraries have also spread all over the country, such as the Guangdong Hong Kong Macao Greater Bay Area Public Library Alliance [4], the six central provinces (Hunan, Hubei, Jiangxi, Anhui, Shanxi and Henan) Public Library Alliance. [5] In addition to the alliance between the university libraries and the public libraries, there are many other types of library alliances, such as the National Library Reference Consulting Alliance [6], the Library Emergency Management Strategic Alliance [7], etc.

From the various types of library alliances mushrooming at home and abroad, all libraries can reach a consensus on the necessity of establishing alliances and join as many library alliances as possible. But why does each province still have its own university library alliance even when they join the national university library alliance like CALIS? Why some libraries which have joined the provincial library alliance also want to join the municipal library alliance? In fact, most of the libraries have joined more than one alliance since each alliance is mainly established to solve a certain kind of problem. However, there are many problems that need to be addressed by establishing the alliances, so they need to join one or the another alliance. Theoretically, the more the participation in the alliances, the greater the benefits libraries reap through joint construction and sharing, but the difficulty lies in cooperating each other. How to build up a library alliance that allows more libraries to join easily and further accelerate the accession of information resources among the member managers? Both the form of library and the form of library alliance are constrained by science and technology over the times. What kind of technical support is necessary for effective library alliance? Presently, two new emerging technologies viz., big data and blockchain can provide new support to library alliances and has tremendous scope to take a step forward.

3. Co-construction of library alliance service platform driven by big data

3.1 Big Data and Library

Big Data refers to a data set that cannot be managed and processed by conventional software tools. Library has a long history of dealing data. In fact, MARC data, which was first produced in the United States in 1961, can be regarded as a kind of big data in the early days. The birth of MARC data has significantly improved the efficiency of library in searching literature resources and laid a foundation for the libraries to move towards integration and automation. In addition to MARC data, each library has accumulated enormous data, reader data, and circulation data in the business management system and are intangible assets of the libraries. Libraries have accumulated massive big data without its full use. This is because these data are scattered in every library and in the absence of any advanced algorithm for processing the data become unusable. Therefore, there is an urgent need to build a new generation of service platform that can collect the big data of all the libraries.

3.2 Feasibility of building a new generation of library service platform

Let's take a look at the statistics of "Double 11" in 2020: from November 1 to 11, 2020, the peak number of orders by Tmall reached 583000 transactions/second. In 2019, there were 3196 public libraries in China, with a total circulation of 90.135 million people, 613.73 million books and literatures to borrow, and 266.09 million people had borrowed books from the libraries [8]. This clearly reflects that only computers can process big data, which is technically feasible for all the libraries across the country to share a common platform for data processing.

The traditional library management system is mainly designed according to the business process of paper books. It is difficult to deal with the diverse library collection resources under the background of the big data that can solve the unified management of all library collection resources. Further, it restricts the process of resource sharing among the libraries and cannot manage the reader's behavior data. In order to meet the needs of business management under the background of big data and to bring a fair level big data technology cooperation, many library alliances are developing new generation library sharing service platforms, which are convenient for the alliance members to carry out big data development and in-depth cooperation on the same platform.

3.3 Construction of library alliance service platform

The service platform of the National Library Alliance can be jointly developed by several institutions. When such platform is available, each library alliance will be integrated, and all the libraries will gradually be absorbed through the convenient joining mechanism to form the service platform of the National Library Alliance.

The construction of the National Library Alliance service platform should organically combine the five elements namely, concept, resources, technology, readers, and mechanism to achieve "innovation from the inside out" [9], so that the platform not only has massive data, but also a management system with multiple advantages such as strong resource integration, new knowledge services, wide coverage, convenient operation, and more user-friendly. The platform should at least include the portal service sub-platform, the shared bibliography retrieval sub-platform, the user monitoring and analysis sub-platform, the information service sub-platform, etc., and each sub-platform has to be connected in series to form a network to provide "one-stop" and "nanny" services (Figure 1).

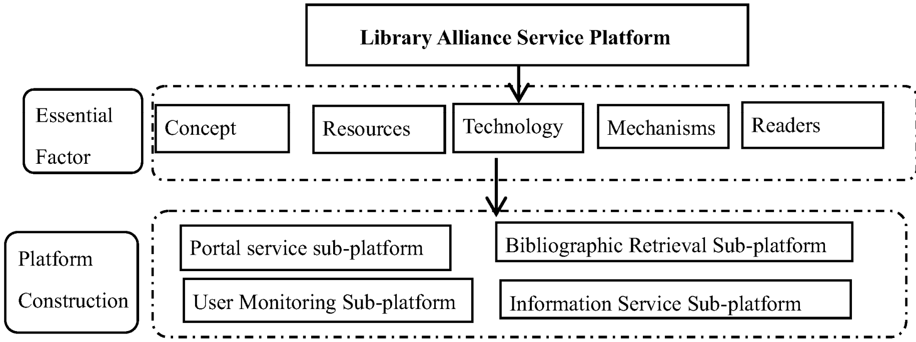


Figure 1. Framework of Library Alliance Service Platform

Although the national shared service platform, which is mandatory for all the libraries, will be created to solve the latest and cutting-edge problems, of which the routine business processing of libraries key module is the early stage development. In fact, a shared bibliographic retrieval platform should first be built based on the data of several major libraries such as the National Library of China, the China Version Library, and the Peking University Library. Each book has only one bibliographic data. Each member library that owns books can add collection information under this bibliography. The collection information can be a private or shared which ensures that the new member libraries can carry out routine business as long as they import their own bibliographic data. After purchasing new books, member libraries can directly store them without cataloging as long as they check duplicates in the general bibliography. This can save a lot of data storage space, cataloging costs and facilitate the centralized data management.

In addition to bibliographic data, the reader data is also the basis for regular library business, and the construction of a nationwide unified user monitoring and analysis platform is also one of the basic works of the National Alliance service platform. As such, it is not difficult to build the reader database. The success of MARC data lies on its specified unified format and allotment of national unified number to each reader. Some information is used as public data whereas, the rest is used as private data with different permissions to different reader data. Different libraries open variety of resources to readers according to their own terms and conditions. Whereas, when one platform is used for circulation of data such as borrowing, repayment, renewal, expiration, and fines, the data is analyzed and managed uniformly, the user tracking activity and interaction data of each library can be synchronized. The artificial intelligence can easily handle enormous volume and complexity of the document information resources especially, for the National Library Alliance service platform. The only solution is to carry out online original text delivery, interlibrary consultation and other services, and build a collaborative service platform for exchange of document information resources. Various strategies like introduction of professional consulting institutions, the establishment of a reference consulting team based on scientific research, university subject librarians and reference consultants of the member libraries of the community, the provision of professional and real-time consulting services for users, and strengthening and expansion of the construction of the collaborative service platform for literature and information resources are in vogue [10].

The construction of the national shared service platform needs to pay attention to the following issues: First, the platform construction is undertaken by the state and is free for all small or large libraries in the country. Second, the platform and data are put into the cloud, which is conducive to the big data processing. Third, attention needs to be provided to data security, storage and backup with different permissions for different member libraries, and protect the data privacy of each library; last but not the least, the platform needs modular development so that different member libraries can choose their own modules according to their specific needs.

The construction of the service platform of the National Library Alliance can save the cost of purchasing, maintenance and upgradation as well as the cost of purchasing the server and storage equipment of each member library. The big data analysis and statistics module can be used to query the borrowing and circulation data, resource guarantee rate and utilization rate of each member library in real time, and even analyze the resource guarantee rate of the member library in the same city, guide the member libraries to coordinate procurement, and improve the efficiency of fund utilization. It can also meet the diversity and personalized needs of the users, ensure social information fairness and bridge the digital divide. Furthermore, new functional modules, such as the module to guide the reading promotion of each library can be developed so that each library can get suitable, scientific and reasonable reading promotion program and develop a module to accurately push the books by finding readers for books and vice versa.

4. Library alliance service platform sharing in the context of blockchain

If free sharing service platform is the reason to attract all the libraries to join the National Library Alliance, then borrowing books from any library in the country is the greatest temptation to attract the readers. The emergence of blockchain technology provides a new imagination space for the library alliance to realize electronic document sharing under the premise of protecting the intellectual property rights.

4.1 Protection of electronic documents in the existing copyright law

With the popularization of smart phones, although paid reading has gradually been accepted by some people, libraries are still the intermediate platform to ensure people's reading and eliminate the data gap. There are many advantages in the dissemination of electronic documents. The number of electronic documents purchased by each library is increasing however, the public service of the library and the copyright protection of electronic documents are the preconditions. The openness of the network makes the electronic documents extremely easy to be copied indefinitely however, the copyright protection is a major obstacle restricting the library's electronic document borrowing service. But, China's Copyright Law and the Regulations on the Protection of the Right of Information Network Communication has exempted the libraries to provide free digital works to the public in their own premises [11]. The copyright law has to be adjusted with time, to balance the intellectual property protection, convenient dissemination of knowledge, expand the scope of use of the library's copyright exception system under the digital reading environment and encourage interlibrary borrowing and document transmission [12].

4.2 New opportunities provided by blockchain technology for electronic document sharing

Blockchain originated from Bitcoin and can be easily applied to library copyright protection. With the gradual expansion of blockchain technology from digital currency to other areas of society, the library community has also carried out in-depth research on the application of blockchain technology and divided into public chain, alliance chain and private chain according to the sharing scope and purpose. The most suitable one for the library alliance is the alliance chain [13]. Compared with the traditional means of copyright and copyright protection, the alliance chain not only has the characteristics of decentralization of the public chain to ensure that data cannot be illegally modified by the organization, but also retains the privacy of the private chain, so that the data storage efficiency is significantly improved. These characteristics provide feasibility in dealing the copyright protection of electronic documents in digital libraries. While publishing electronic documents, publishers generate unique alliance chain for each electronic document after encryption, and then generate a corresponding number of sub chains according to the number of electronic documents issued. It is safe and reliable to manage copyright through alliance chain, since it is difficult to tamper. Publishers can sell the sub chain of the alliance chain to the library distributor or directly to the library. Purchasing a sub chain is equivalent to purchasing an electronic document. The data on the sub chain can be set to the reader online reading mode or download reading mode according to the management needs, and the reader's query however, the downloads and other operations will leave traces on the chain [14]. The sub chain is equivalent to a powerful book borrowing card, which can objectively and truly record the borrowing amount, download amount, appointment information, etc. of each electronic document.

4.3 Construction of reading sharing module of the National Library Alliance service platform

The biggest advantage of the application of the alliance chain technology to the library alliance is that it can limit the random copying of electronic documents and managed completely as paper documents. When electronic documents have the non-random duplication of paper-based documents, they should have the same circulation authority as paper-based documents from a legal perspective. As library books are borrowed for home reading, the library alliances can lend and return paper-based documents similarly, electronic documents can also be issued to the readers and returned.

The National Alliance Service Platform needs to develop a national shared e-book reading module (hereinafter referred to as the "reading module"). Readers who borrow a sub chain from the library have the right to read the book. They can only read the book directly on the reading module, but cannot download or copy it to protect intellectual property rights. For books within the copyright protection period, the reading module can give each library a corresponding number of sub chains based on the number of paper books owned by each library through the alliance chain technology. The current law in China does not specify how many electronic versions can be owned by a library for the digital collection of paper books, but at least one should be owned. The sub chains owned by each library are one-to-one corresponding to the paper books. The books can even be borrowed in the library premises of the entire library consortium (the agreement of joining the consortium can stipulate the interlibrary borrowing terms of electronic books). The copyright law neither explicitly supports interlibrary borrowing nor prohibits it. In reality, the interlibrary document transmission is quite common. The construction of reading modules is highly shared and avoid repeated construction of each library. The big data generated in the borrowing process is stored in the alliance chains and sub chains, which is also convenient for future analysis and processing.

How to ensure that the number of sub chains of each alliance chain on the reading module is enough for the readers to read? First, the system is set to allow readers to borrow when they open the documents, and return when they close the documents. The maximum borrowing period is not more than 24 hours, so that the book borrowing efficiency is maximized. When the borrowing period of each library is 30 days the book borrowing efficiency increases by 30 times so, the individual library savings are obvious. Second, the documents are timely returned and the borrowing volume generally decreases with the passage of time. Therefore, depending on the number of sub chains converted from the paper books owned by each library, most of the national reading needs get fulfilled. Especially, for new bestseller books, the number of sub chains need to be increased. For e-books on the reading module, the system can calculate the recommended purchase index of books according to the readers demand and the scores of the readers. Each member library will purchase books according to its own purchase budget and combine with the recommended purchase index to supplement the number of sub chains of the platform. Third, libraries with large borrowing volume and insufficient existing sub-chains can appropriately increase the purchase quantity or mobilize social forces to donate.

5. Conclusion

Big data has provided new impetus in the construction of the National Library Alliance service platform and ensured close cooperation in library alliance and high sharing of resources and data. The more data the platform has, the better the advantages of the algorithm can be brought into full play. Relying on the support of the National Library Alliance service platform, better library services can be offered not only within the country but even at the global level. Big data and blockchain technology drive library changes and in turn promote technological progress. Through this series of virtuous cycles, we can create a win-win situation of technology, library reform and transformation.

References

[1] Zhao Naixuan, Jin Jieqin, Feng Qin. Empirical Study on the Construction and Development of China's Cross system Regional Library Alliance [M]. Nanjing: Nanjing University Press, 2018.

[2] Liu Kuibo. Research on Co construction and Sharing of Characteristic Databases of Regional University Library Alliances [J]. Information Exploration, 2018 (11): 112-116.

[3] Wang Yi, Shao Bo. Development Status of Domestic University Digital Library Alliance: Lack of Connotation and Restructuring [J]. Library Science Research, 2020 (4): 87-93.

[4] Teng Xue, Deng Min, Yan Ping. Research on Library Collaboration Alliance Mechanism in the Context of Guangdong Hong Kong Macao Greater Bay Area [J]. Library Research and Work, 2020 (9): 39-42.

[5] Hunan Hubei Jiangxi Anhui Public Library Alliance. Introduction to the Public Library Alliance of the Six Central Provinces (Hunan Hubei Jiangxi Anhui Shanxi Henan) [EB/OL]. (2020-9-27) [2022-11-18] http://www1.library.hb.cn/gaikuang.html.

[6] Xie Chaohui. Analysis on User Demand Characteristics and Service Status of National Library Reference Consulting Alliance [J]. Library Forum, 2015 (2): 48-54.
[7] Guo Xiao. The Operating Characteristics and Enlightenment of Foreign Library Emergency Management Strategic Alliance -- Also on the Construction of China's Library Emergency Management Alliance [J]. Library Work and Research, 2020 (7): 58-63.
[8] Website of the Ministry of Culture and Tourism. Statistical Bulletin of the Ministry of Culture and Tourism of the People’s Republic of China on Culture and Tourism Development in 2019 [EB/OL]. (2020-06-22) [2022-11-18] http://www.gov.cn/xinwen/2020-06/22/content_5520984.htm.
[9] Gan Lu. On Library Immunity [J]. Journal of University Library, 2016 (6): 100-104
[10] Li Qiurong. Challenge and Reconstruction of Library Copyright Exception System in the Age of Big Data [J]. Library Work and Research, 2019 (7): 10-14, 49.
[11] Xu Lihua, Nie Feixia. Construction of information resource security sharing model based on digital library alliance chain [J]. Library and Information Work, 2020 (3): 53-58.
[12] Li Jinyang. The Combination of Library and Alliance Chain: Concept, Demand and Prospect [J]. New Century Library, 2020 (8): 5-10.
[13] Xu Lihua, Nie Feixia. Construction of information resource security sharing model based on digital library alliance chain [J]. Library and Information Work, 2020 (3): 53-58.
[14] Li Jinyang. Combination of library and alliance chain: concept, demand and prospect [J]. New Century Library, 2020 (8): 5-10.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230173

According the Date to Investigate and Analysis on the Current Situation of Preparations in Medical Institutions in Chongqing Based on Grounded Theory

Yanying Yang[1], Xueqin Hu, Bingqian Zhang, Xiaohong Zheng and Juan Xu[2]
Chongqing Medical and Pharmaceutical College, Chongqing, China.

Abstract. Objective: Investigate the development status of preparations in medical institutions in Chongqing, analyse the key factors restricting the development of preparations in medical institutions in Chongqing, and put forward targeted countermeasures and suggestions, so as to promote the sustainable and healthy development of preparations in medical institutions in Chongqing. Methods: Data were collected by means of literature review, expert interview, questionnaire surveys (Questionnaire Star, which is a platform providing functions equivalent to Amazon Mechanical Turk). SPSSAU was used for statistical analysis, and the scientific research method of classic Grounded Theory was used for data processing and theoretical construction[1]. The in-depth situational grounded research was conducted on the preparations of medical institutions in Chongqing. Results and conclusion: Due to the insufficient attention paid by the government, hospitals and the market, the development of preparations in Chongqing's medical institutions could not keep up with the development of the pharmaceutical industry and drug administration supervision. In view of this, the government should coordinate the deployment of "capital increase, staff increase, strength increase " and "integration and joint construction" to promote development, focus on the research and development of Chinese medicine preparation varieties that are in short supply in the market and are urgently needed clinically, and incubate new drugs in the process of inheritance and innovation. The preliminary theoretical summary of the current situation and development of pharmaceutical preparations in medical institutions in Chongqing can provide a basis for decision-making by management departments and promote its sustained and healthy development.

Keywords. Grounded Theory, Medical Institutional Preparation Chongqing, SPSSAU.

Corresponding author: Juan Xu, Chongqing Medical and Pharmaceutical College. Email: 541049385@qq.com; Xiaohong Zheng, Chongqing Medical and Pharmaceutical College. Email: 446091614@qq.com
Fund project: Decision Consultation and Management Innovation Project in Shapingba District, (Jcd202104), Research Project of Chongqing Medical and Pharmaceutical College (ygz2019102, CQYGZJG20). First author: Yanyin Yang, Chongqing Medical and Pharmaceutical College. Email: 20079@cqmpc.edu; 383765857@qq.com

1. Introduction

The Grounded theory methodology (Grounded theory for short) is an independent research methodology, which was founded by American sociologists Barney G Glaser and Anselm L Strauss in 1967, and has been developed and deepened since then. It plays a foundational role in qualitative research and is known as the precursor of the "qualitative revolution" [2]. In the historical evolution of nearly half a century, Grounded theory researchers have been constantly studying, improving and developing, making Grounded theory finally form three versions that are both interrelated and not completely identical: the original version of Grounded theory initially elaborated by Glaser and Strauss classical Grounded theory; Strauss and Corbin's procedural Grounded theory, and Charmaz's constructive Grounded theory. The establishment and development of the methodology of Grounded theory is a major breakthrough in qualitative research [3]. As Strauss pointed out in his works published in 1987 and 1990, Grounded theory research is a continuous cycle process of collecting data and testing at the same time, which has contained testing [4]. Therefore, "from now on, almost all qualitative research will be applied to the basic principles or specific operating procedures of Grounded theory" [5]. Grounded theory has been widely used in sociology, anthropology, psychology, pedagogy and other fields abroad and has achieved fruitful results, while the domestic literature on Grounded theory is relatively small. In recent years, the domestic application of Grounded theory in education, business, management and other fields has increased year by year, and a large number of scholars have widely applied it to various fields such as medicine and health, Chinese medicine culture, and even the construction of Chinese medicine theory, and made great progress. Among the specific disciplines of pharmaceutical management, Yang Yue[6] first mentioned in the "Research Methodology of Pharmaceutical Management", and then the author applied it to specific drug registration, adverse drug reaction reporting and monitoring studies in Chongqing [7]. This article then conducts a grounded study on the development of pharmaceutical preparations in medical institutions in Chongqing. The specific research procedure is shown in Figure 1.

The preparations of medical institutions are directly connected with the clinic, which can be adjusted in time to meet the clinical needs. The circulation loss cost is low, and there are many characteristic preparations with exact curative effects, which have irreplaceable outstanding advantages[8]. With the development of the pharmaceutical industry, many hospital preparations that are clinically practical have been replaced by commercially available drugs. In addition, the preparation conditions in the preparation room of medical institutions cannot meet the requirements of current drug regulatory regulations[9], the production mode of medical institutions has limitations, and medical institutions have no pricing power for preparations[10]. In recent years, the number of types of preparations of medical institutions in our city, as well as the perennial varieties and quantities of preparations, has shown a significant downward trend, The number of medical institutions that have the capacity to make preparations and have obtained the Dispensing License for Medical Institutions has decreased year by year, and some medical institutions have closed their preparation rooms. The suspension and cancellation of preparations of characteristic medical institutions with good clinical effects and good reputation are a loss for patients and the pharmaceutical industry. How to find a way out for the development of pharmaceutical preparations in medical institutions under the shrinking trend is worth thinking deeply by pharmaceutical practitioners and relevant departments in medical institutions. In view of this, it is very necessary to carry out research on the current situation and policies of preparations in medical institutions in Chongqing and give corresponding suggestions.

2. Respondents and methods of survey

2.1 Respondents of survey

In this study, a cross-sectional questionnaire was used[11]. The respondents of survey were the ones in charge of the pharmaceutical department (pharmacy) or preparation room of a medical institution in Chongqing with the approval number of medical institutional preparations, who was willing to cooperate with this survey. The questionnaire consists of two types of questions: closed questions and open questions. single choice questions and multiple-choice questions are used to understand the basic situation of preparations in medical institutions in Chongqing; Open questions are used to collect the opinions and suggestions of investigators on the development of preparations in medical institutions in Chongqing. The questionnaire mainly relies on the "Questionnaire Star" electronic questionnaire platform to fill in and collect data online. A small amount is recycled through WeChat or email.

2.2. Methods of survey

2.2.1 Questionnaire content design.

The author discussed the design repeatedly with experts in drug supervision and management, and conducted a questionnaire survey on the selected subjects. The main contents of the questionnaire include: the basic information of the interviewees and their institutions, the status and problems of medical institutions' preparations, and the opinions and suggestions on the development of medical institutions' preparations.

2.2.2 Quality Control.

After the questionnaire was designed, the opinions of experts were collected to gradually optimize the questionnaire. After three rounds of discussion, review and correction by experts in drug supervision and management, this formal questionnaire has been finalized to ensure the rationality and feasibility of the content design. After the questionnaire was collected, check and accept the logicality, integrity and accuracy of the questionnaire one by one.

2.3. Data entry and statistical analysis

Two entries were used for comparison, contrast and errors correction. The basic information of the collected data is expressed as a percentage (%), and the qualitative analysis of the string variable data is conducted based on the grounded theory coding. The influencing factors were statistically analysed by SPSSAU.

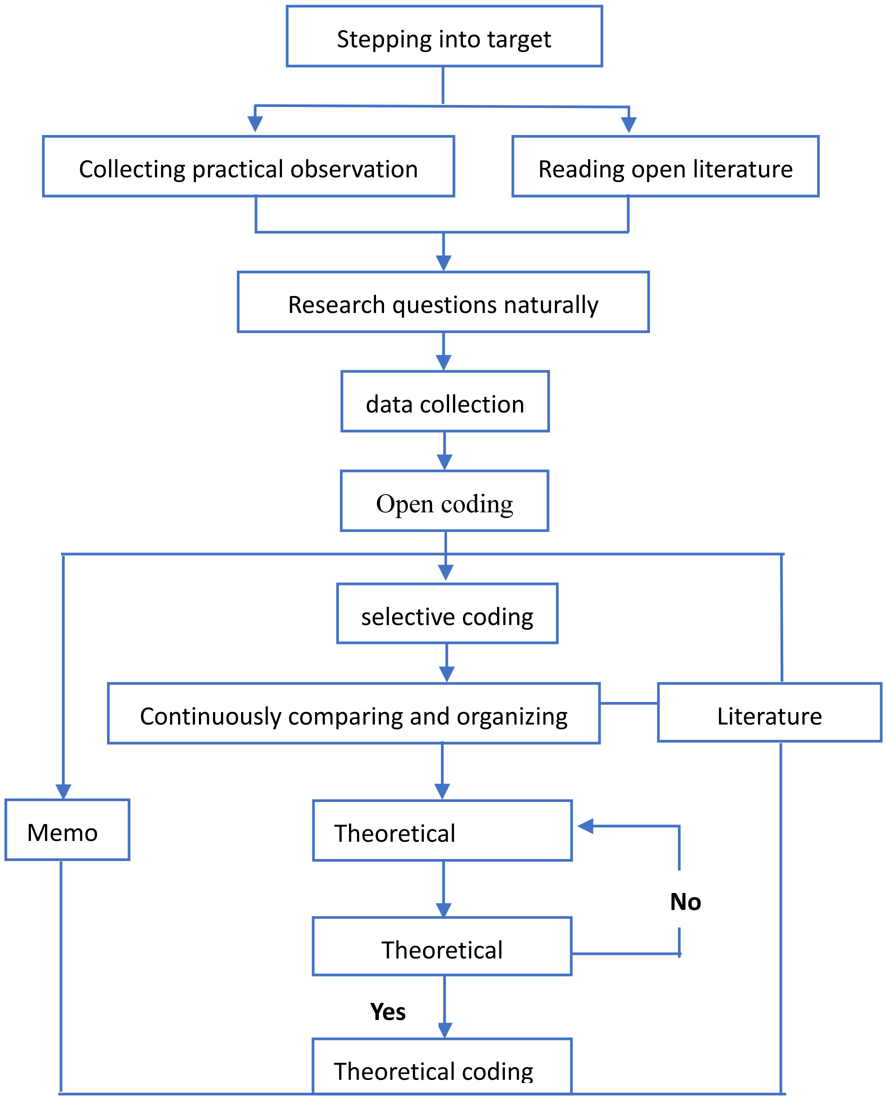


Figure 1: Research Program Diagram

3. Results and Analysis

3.1. Basic information

This survey questionnaire was distributed to 19 medical institutions holding the "Medical Institution Preparation License" in 2021, and the questionnaire code was shared in the medical institution management contact group. Finally, a total of 38 questionnaires were collected, involving 27 medical institutions. 20 expert interview papers were distributed and 18 were recycled, with a recycling rate of 90%. If there is inconsistent data among different personnel of the same medical institution in the collected questionnaire, please contact the relevant person in charge and regulatory authorities by phone to verify and confirm the correct and valid data, while excluding invalid data. The opinions collected from the open questions and the expert interview volumes are all included in the Grounded theory data coding analysis. The 27 medical institutions participating in the survey are distributed in 20 districts (counties) of Chongqing, including 9 Class II Class A hospitals (33.33%), 1 Class III hospital (3.70%), and 17 Class III Class A hospitals (62.96%). 19 hospitals with more than 500 employees accounted for 70.37%, their pharmacy staff accounted for 3.00-6.00%, 2 hospitals with 100-500 employees accounted for 7.41%, their pharmacy staff accounted for 6.00% and 8.50%, 2 hospitals with 50-100 employees accounted for 7.41%, their pharmacy staff accounted for more than 80. 00%, 4 hospitals with less than 50 employees accounted for 14.81%, and their pharmacy staff accounted for 4.10-6.00%. Half of the preparers was the person in charge (director and deputy director) of the preparation room (center), and the rest were pharmacists or Chinese pharmacists excepts for two drug control personnel. The 18 experts were interviewed included experts from municipal drug supervision and administration institutions, leaders of medical schools and directors of teaching and research departments, heads (directors and deputy directors) of hospital preparation departments (centers), and professionals with years of experience in preparation related work in medical institutions.

Among the 27 medical institutions participating in the survey, 8 (29.63%) medical institutions set up preparation rooms and held 255 preparation approval documents with the Medical Institution Preparation License, including 183 chemical preparations which accounted for 71.76%, and 72 traditional Chinese medicine preparations which accounted for 28.24%. The approval number is mainly concentrated in Chongqing Traditional Chinese Medicine Hospital, which holds 153, accounting for 60.00%, and the other 7 hospitals hold several to twenty or thirty respectively. In terms of personnel allocation in the preparation room, 35 people in Chongqing Hospital of Traditional Chinese Medicine, the remaining 7 hospitals have 3-9 people. Among them, 25 Chongqing Hospital of Traditional Chinese Medicine are engaged in preparation in hospital, the remaining 1 to 3 people are engaged in preparation inspection in hospital, 1 to 3 people are engaged in preparation management in hospital, 7 people from Chongqing Hospital of Traditional Chinese Medicine are engaged in preparation research and development in hospital, 3 medical institutions have no R&D personnel, and the other 4 hospitals have 2-3 people. 8 medical institutions have 255 approval numbers, including 3 tablets, 12 capsules, 4 pills, 2 granules, 0 injections, 28 mixtures, 8 oral liquids and 7 transdermal preparations; No special preparation for children; Only one patent right is granted; In the past five years, 4 approval numbers were added and 96 were cancelled. Of the 255 documents, 112 were prepared by our hospital, accounting for 43.92%, 2 were entrusted, accounting for 0.78%, and 141 were not prepared, accounting for 55.29%.

Of the 27 medical institutions participating in the survey, only 7 (25.93%) hold registration numbers for the preparation of traditional Chinese medicine preparations using traditional techniques, including 2 capsules, 5 pills, 11 granules, 43 mixtures, 1 transdermal preparation and 34 other dosage forms, totaling 96. No special preparation for children. Only one patent right is granted. In recent five years, 29 new filing numbers were added. The number of preparations in hospital is 48 which accounted for

50.00%. The number of entrusted preparations is 19, accounting for 19.79%. The number of unprepared preparations is 13, accounting for 13.54%.

Among the 27 medical institutions participating in the survey, 1 (3.70%) has 10-20 classic famous prescriptions or famous Chinese medicine proven prescriptions that have been used clinically for more than 5 years. 1 (3.70%) has 5-10 classic famous prescriptions or famous Chinese medicine. 2 (7.41%) has 3-5 classic famous prescriptions or famous Chinese medicine, 5 (18.52%) has 1-3 classic famous prescriptions or famous Chinese medicine. 18 (66.67%) has no classic famous prescriptions or famous Chinese medicine.

Of the 38 respondents, 20 (52.63%) thought it was necessary to further improve the laws and regulations related to preparations in medical institutions, 12 (31.58%) thought it was unnecessary, and 5 (13.16%) thought it was indifferent. 18 people (47.37%) had plans to develop or apply for new medical institutions' preparations, while 20 people (52.63%) had no plans. However, as for whether they are interested in further developing preparations from medical institutions into new drugs on the market, 22 people (57.89%) were very interested in it, while only 3 people (7.89%) were not interested at all, and 13 people (34.21%) 13 people (34.21%) followed the instructions of the leaders.

3.2. Analysis of factors affecting the development of preparations in medical institutions

SPSSAU was used to analyze the factors affecting the development of pharmaceutical preparations in medical institutions. The KMO test value was 0.666 (>0.5); Bartlett sphericity test ($p<0.05$) is shown in table 1, indicating that the research data is suitable for principal component analysis. According to the general principle of extracting the number of principal components (the cumulative variance contribution rate of the extracted principal components is ≥ 85%), the information that the first five principal components can explain 87.43% of the total variance of the original variables is shown in table 2. Although the characteristic roots of the fourth and fifth principal components are lower than 1, it is appropriate to extract the first five principal components in order to fully consider the relevant factors affecting the development of medical institutions as far as possible. See table 3 for load factor table.

Table 1. KMO and Bartlett's tests.

KMO value		KMO value
	Bartlett's Sphericity Test	88.558
Bartlett's Sphericity Test	*df*	28
	P value	0.000

Table 2. Variance interpretation rate table.

Number	Characteristic root			Principal component extraction		
	Characteristic root	Variance interpretation rate%	Cumulative %	Characteristic root	Variance interpretation rate%	Cumulative %
1	3.232	40.404	40.404	3.232	40.404	40.404
2	1.31	16.378	56.782	1.31	16.378	56.782
3	1.061	13.269	70.051	1.061	13.269	70.051
4	0.777	9.718	79.77	0.777	9.718	79.77
5	0.613	7.659	87.428	0.613	7.659	87.428
6	0.439	5.492	92.92	-	-	-
7	0.351	4.39	97.31	-	-	-
8	0.215	2.69	100	-	-	-

Table 3. Load factor table.

Name	Load factor					Commonality (Common factor variance)
	principal component 1	principal component 2	principal component 3	principal component 4	principal component 5	
Raw material source quality assurance and cost	0.74	-0.16	0.244	0.354	-0.06	0.762
Product R&D technology	0.668	-0.376	-0.132	0.115	-0.498	0.866
Product pricing and revenue	0.608	-0.59	0.205	0.004	0.31	0.856
Clinical efficacy and adverse reactions of the product	0.805	-0.063	-0.198	-0.031	0.413	0.862
Preparation and inspection equipment and technology	0.763	0.217	0.119	-0.369	-0.291	0.864
Allocation of pharmaceutical technicians	0.608	0.333	-0.576	-0.327	0.072	0.924
Importance of hospital leaders	0.34	0.465	0.737	-0.203	0.064	0.92
Support from the hospital and relevant municipal departments	0.393	0.645	-0.122	0.595	0.033	0.94

3.3. Problems that need to be supported in pharmaceutical R&D planning of medical institutions

SPSSAU was used to conduct principal component analysis on the problems that need to be supported by the pharmaceutical R&D planning of medical institutions, and the KMO test value was 0.704 (>0.5); Bartlett sphericity test ($p<0.05$) (table 4) shows that the research data is suitable for principal component analysis. According to the general principle of extracting the number of principal components (the cumulative variance contribution rate of the extracted principal components is ≥ 85%), the first four principal components can explain 90.13% of the total variance of the original variables, and the first three principal components can explain 82.51% of the total variance of the original variables. See table 5. The characteristic roots of the third and fourth principal components are lower than 1, and the absolute value of the load coefficient of the fourth principal component is less than 0.4. It is appropriate to comprehensively consider the first three principal components for medical extraction. See table 6 for load factor table.

Table 4. KMO and Bartlett's tests.

KMO value		0.704
Bartlett's Sphericity Test	Approximate chi square	30.274
	df	15
	P value	0.011

Table 5. Variance interpretation rate table.

Number	Characteristic root			Principal component extraction		
	Characteristic root	Variance interpretation rate %	Cumulative %	Characteristic root	Variance interpretation rate %	Cumulative %
1	2.947	49.121	49.121	2.947	49.121	49.121
2	1.359	22.655	71.776	1.359	22.655	71.776
3	0.644	10.736	82.512	0.644	10.736	82.512
4	0.457	7.614	90.126	0.457	7.614	90.126
5	0.348	5.793	95.919	-	-	-
6	0.245	4.081	100	-	-	-

Table 6. Lad factor table.

Name	Load factor				Commonality (Common factor variance)
	principal component 1	principal component 2	principal component 3	principal component 4	
Need financial support	0.596	-0.646	0.118	0.351	0.91
Need registration policy consultation and guidance	0.785	0.042	-0.396	-0.395	0.931
Need R&D technical support	0.727	-0.493	0.087	-0.167	0.807
Need to commission production	0.7	0.361	-0.449	0.378	0.965
Need independent pricing	0.496	0.745	0.345	0.044	0.922
Necessity to relax the regulation restrictions and expand the clinical use to increase the benefits	0.843	0.106	0.381	-0.075	0.873

3.4.Rooting analysis of open issues

3.4.1 Problem posing. Consult extensively the policy documents related to drugs, drug management, drug registration and hospital preparations, be familiar with the hospital preparations registration management and other work, carefully observe the preparation situation in the work, and observe the interaction between the subject and object. Check the registration application materials. By comparing and analyzing the practical data with the literature reviewed, we determined to carry out the relevant research on hospital preparations.

3.4.2 Data collection. According to the sorting and analysis of the mastered data, the expert interview outline was initially prepared. In consideration of the comprehensiveness and authority of the relevant information, the interviewees were determined to be the experts in drug supervision and management, the person in charge of preparations in the hospital, the person in charge of preparations in the medical institutions of the second and third level hospitals in Chongqing, and the college medical pharmacy experts. The interview outline and interview quality shall be supervised by the expert in drug supervision and management. Please select the interviewees to fill in the expert interview outline and summarize all the collected data for analysis and processing.

3.4.3 Coding. According to the qualitative research method of grounded theory, the new concepts and measures proposed or redefined are analyzed and compared with the previous literature reports for the collected data layer by layer comparative analysis and progressive coding, open coding, selective coding and theoretical coding [7].
Open coding Considering the inconvenient source of the analysis tool software, which cannot replace the human brain analysis, and there are not too many text materials, this study uses manual coding to open the data obtained, and strives to maintain a completely free and open state in the process, line by line, word by sentence, and strive to be close to the original meaning, finally breaking the data on the development status and direction of the preparations in Chongqing medical institutions into 68 concepts, The proposed data on the development of preparations in medical institutions in Chongqing were broken into 141 concepts. Through the comparative analysis between concepts, identify every valuable phenomenon and recombine relevant concepts into categories. Finally, 14 and 15 categories were extracted respectively. In view of space limitation, the concept category, category and concept category constituting category are omitted.
Selective coding The 14 categories obtained from the data on the development status of preparations in medical institutions in Chongqing and the 15 categories obtained from the data on the suggestions on the development of preparations in medical institutions in Chongqing were analyzed in depth one by one. The data were reviewed and the relevant literature was constantly consulted for comparison. The two core categories were obtained and further correlated and selectively coded using the "canonical model". At this stage of the study, nearly 100 new related literatures were retrieved. Through detailed comparative analysis, the category characteristics obtained can no longer be further developed, and it was judged as theoretical saturation. See table 7-8 for the paradigm table.

Table 7. The paradigm model about core categories.

Causal condition			**Intermediary conditions**	**Action/ interaction strategy**	**Intermediary conditions**	**Phenomenon**	**result**	core category
A1	A7, A2, A9, A13	A5, A6						AA1: The government, hospitals and the market did not pay enough attention, and the preparations of medical institutions could not keep up with the development of the market, the pharmaceutical industry and the drug administration
Market	production development	Drug administration supervision	A12	A14	A10	A4, A8	A3	
AA1 context: In the market dimension, the preparations of medical institutions are similar to or substitute for a large number of products. In the production and development dimension, the development of preparations of medical institutions and their productivity can not keep up with the development of modern pharmaceutical industry [8]. In the drug administration and supervision dimension, the supervision of the drug administration on modern pharmaceutical industry is applied to the preparations of medical institutions, which makes them at a loss. The government, hospitals and the market do not pay enough attention to and invest in the preparations of medical institutions, and the preparations of medical institutions shrink rapidly, The existing preparations are only distributed in the top three hospitals in the main city, and traditional Chinese medicine preparations are far from new drugs, not to mention the role of "incubator" for new drug research and development.								

Table 8. The paradigm model about core categories.

Causal condition	Intermediary conditions	Action/interaction strategy			Intermediary conditions	Phenomenon		result	core category
		A2, A7, A9, A15	A10	A5, A12	A6, A11				
A3, A4	A1	The existing preparation products are "optimized"	Promoting the filing of traditional Chinese medicine preparations	"Increase capital, increase staff, increase strength", "integrate and jointly build", develop medical institutions, prepare and incubate new drugs	The government coordinated the deployment, improved and supported a series of policies, systems, specifications and technical documents, and put them into place to promote the development of new preparations and incubate new drugs in hospitals	A13	A14	A8	AA2: The existing pharmaceutical products "select the best to improve the quality", the clinical urgent need, diversified and characteristic traditional Chinese medicine preparations are the focus, and "increase capital and personnel to increase strength" and "integration and joint construction" promote development.
AA2 context: Those who have advantages in the pharmaceutical incubation of new drugs in medical institutions believe that new drugs can be marketed only with funds and transformation platforms. The development of preparations is difficult, and there is a long way to go. The development of preparations in medical institutions depends to some extent on the degree of attention paid by the hospital leaders. The targeted policy is to "select the best and improve the quality" of existing preparations; The government coordinated the deployment, improved and supported a series of policies, systems, specifications and technical documents, and promoted and implemented them in place, "Increase capital, increase staff and increase strength" and "integrate and build together" promote the development of new preparations and incubate new drugs in hospitals. Focus on the research and development of the varieties of traditional Chinese medicine preparations that are in short supply in the market and urgently needed in the clinic, and develop in a diversified and characteristic way in the process of inheritance and innovation.									

In view of space limitations, the verification of AA1 and AA2 evidence chains will not be narrated here. Through comparative analysis and coding of collected data layer by layer and comparison with previous literature reports, three new concepts are proposed or redefined: "selecting the best and improving the quality", "increasing capital, increasing staff and increasing strength" and "integrating and building together".

"Selecting the best and improving the quality" is to adapt to the high standard requirements of the new situation on the safety, effectiveness, quality control and other aspects of medical institutions' preparations. Guided by clinical needs and efficacy, the hospital, in combination with the actual situation, conducts a careful and responsible analysis of the existing preparation varieties, specifications, output, clinical data, etc., screens and makes decisions, and selects the preparation varieties with clear clinical positioning and obvious clinical value, Concentrate on increasing investment, refer to the latest requirements of pharmaceutical preparations and new drug research and development in medical institutions, and actively carry out intensive research in various aspects such as optimizing processes, improving standards and accurately enhancing clinical safety and effectiveness with new technologies, so as to obtain the development of pharmaceutical preparations in medical institutions and the transformation of new drugs.

"Increasing capital, increasing staff and increasing strength" is to increase capital investment, strengthen the construction of R&D team, improve professional and technical literacy, enhance R&D innovation, and develop high-quality traditional Chinese medicine preparations with clear positioning and significant effects into new drugs on the market through inheritance and innovation.

"Integrating and building together" means that there are many difficulties for hospitals to develop medical institutions to incubate and market new drugs on their own. The government can uniformly coordinate and deploy the featured preparations that hospitals focus on, improve and support a series of policies, systems, specifications, processes and technical documents, promote and implement them in place, integrate preparation R&D and preparation resources, and build a preparation and achievement transformation center through "medical school (research) enterprise" collaboration, We will promote the development of pharmaceutical preparations and new drug incubations in medical institutions, while avoiding the loss of state-owned assets.

Theoretical coding Through further investigation of the two core categories and interactive comparison with the original data, we can find the implicit interrelationship between concepts or categories, sort out and organize the substantive coding concepts and categories with the help of the "story line" of the development and research questions, and initially build a theoretical framework, that is :Due to the insufficient attention paid by the government, hospitals and the market, the development of preparations in Chongqing's medical institutions could not keep up with the development of the pharmaceutical industry and drug administration supervision. In view of this, the government should coordinate the deployment of "capital increase, staff increase, strength increase" and "integration and joint construction" to promote development, focus on the research and development of Chinese medicine preparation varieties that are in short supply in the market and are urgently needed clinically, and incubate new drugs in the process of inheritance and innovation.

We further searched the literature and made a comprehensive review and comparative analysis. It is especially suggested that the data information of the literature retrieved is nationwide rather than limited to Chongqing. The "story line" of the development status and development suggestions of Chongqing medical institution preparations developed by the grounded theory method in this study basically covers the existing problems of literature disclosure and suggestions for solutions.

4. Discussion

4.1. The medical institutions in Chongqing have a weak foundation of preparations, and the Traditional Chinese Medicine Hospitals in Chongqing account for half of the whole city.

Among the hospitals participating in the survey, the 8 hospitals with preparation rooms are equipped with 3-9 staff members related to preparations from medical institutions, except Chongqing Traditional Chinese Medicine Hospital, which has 35 staff members. The 27 medical institutions participating in the survey held 255 drug preparation approval numbers in total, of which 153 were held by Chongqing Hospital of Traditional Chinese Medicine, accounting for 60.00%. The new approval numbers in recent 5 years are all newly added by Chongqing Traditional Chinese Medicine Hospital. There are total 112 preparations make up by themselves medical institution and 65 (58.04%) come from Chongqing Traditional Chinese Medicine hospitals. The 27 medical institutions participating in the survey have 96 filing numbers of traditional Chinese medicine preparations, 42 of which are held by Chongqing Hospital of Traditional Chinese Medicine, accounting for 43.75%. The total number of preparations prepared by this hospital is 48, 25 of which are held by Chongqing Hospital of Traditional Chinese Medicine, accounting for 52.08%. There are 35 classic famous prescriptions or 5-10 proven traditional Chinese medicine prescriptions that have been used clinically for more than 5 years

4.2. Main factors affecting the development of pharmaceutical preparations in medical institutions

The five main factors affecting the development of preparations in medical institutions are in turn preparation, testing equipment and technology; Product pricing and revenue; Degree of attention paid by hospital leaders; Support from the hospital and relevant municipal departments; Clinical efficacy and adverse reactions of the product. The analysis results are basically consistent with those involved in the current development of preparations in medical institutions and the collection of suggestions on open issues.

4.3. Hospitals with preparation R&D plans of medical institutions mainly need three aspects of support

Hospitals with pharmaceutical R&D plans of medical institutions mainly need support from three aspects, namely, expanding clinical use to increase income; Independent pricing is required; Entrusted production is required. Relax the restriction on dispensing and expand the clinical use. On the one hand, hospitals can increase the income, on the other hand, they can collect more clinical data to prepare for the incubation of new drugs. It is worth noting that in the process of expanding the clinical use, the safety and effectiveness of products should be analysed accurately with reference to the requirements of clinical research materials of new drugs. Entrusted production allows pharmaceutical enterprises with advantages in production conditions and resources to easily solve problems such as the hospital's own allocation of no factory, no equipment and no technicians, and prompts the urgent need to solve the construction of hospital enterprise communication platform.

4.4. Problems of new drugs incubated by medical institutions

There are a total of 8 options for the question of which aspects of support are needed to further research and development of medical institutions' formulations into new drugs for the market,among the 38 valid questionnaires, none of them chose "no support is needed, the hospital can solve the problem by itself", while 35 (92.11%) chose "technical guidance for research and development" and "guidance for policy interpretation on requirements related to registration and approval"; 19 people (50.00%) chose to lower the standard for registration review, while other options exceeded 70%. This problem is not suitable for principal component analysis or factor analysis. In addition to the unscientific setting of exclusion options, it also shows that the transformation from hospital preparations to new drugs is very technical, and the requirements for the registration and approval of new drugs make hospital personnel confused. From medical institutions to new drugs on the market, it is also the same as new drug research and development, with high investment and long cycle. We really need to solve many problems.

5. Conclusion

After reviewing the problems and suggestions of the open description, the grounded research summarized that: due to insufficient attention from the government, hospitals and the market, the development of preparations in Chongqing's medical institutions could not keep up with the development of the pharmaceutical industry and drug administration. In view of this, the government should coordinate, deploy, improve and support a series of policies, systems, specifications and technical documents to promote the development of medical institutions' preparations by "selecting the best and improving the quality" of existing medical institutions' preparations, "increasing capital and personnel to increase strength", "integrating and co constructing", and focus on the research and development of Chinese medicine preparations that are in short supply in the market and urgently needed clinically, and incubate new drugs through inheritance and innovation.

6. Deficiencies and prospects

There are many and complicated problems in pharmaceutical development and new drug research and development in medical institutions. When designing the questionnaire items, the length of the questionnaire was taken into account, so the design of homogeneous problems was not enough and the correlation was not strong enough. If further classification and separate volume survey were carried out, more accurate results could be obtained.

The scientific research method of Classic Grounded Theory is used to discuss the development of preparations in medical institutions in Chongqing. However, due to the limitation of time, energy and research ability, this topic has many shortcomings. For example, in the research of preparations in medical institutions, many important issues such as process standards and application materials have not been further classified and deeply rooted, the development of preparations in medical institutions has summarized that the development direction is the clinical urgent need for characteristic traditional Chinese medicine preparations, but the specific way to carry out the research and development of the clinical urgent need for characteristic traditional Chinese medicine preparations has not been further discussed, which needs to be further deepened in the future research. The whole article is based on the data collection of questions and suggestions, and the preliminary theoretical summary proposed needs to be verified and improved in practice. At the same time, the study failed to conduct research on more provincial and municipal related issues. In the future, more samples need to be collected for research, and compared with the data obtained in this study, to continue to deepen and enrich the connotation of the concepts and categories proposed in this topic.

Acknowledgments

Special thanks to the leaders and all members who participated in and supported the questionnaire survey on the status of preparations in medical institutions in Chongqing. The results of the study come from their joint efforts.

References

[1] JiaXudong,Tan Xinhui.The Actual Value of the Classical Grounded Theory and Its Spirit to China Management Research [J]. Chinese Journal of Management, 2010,7(5): 656-665.
[2] Denzin N K, Lincoln Y S. Handbook of qualitative research[M]. Sage Publications, Inc, 1994.
[3] Zhang Mengzhong, Mark Huozhe (American). Introduction to qualitative research Methodology [J]. China Administration. 2001, (11): 39-42.
[4] Strauss A, Corbin J, and Xu Zongguo (translated). Introduction to Qualitative Research [M]. Taipei: Juliu Book Company, 1997.
[5] Feng Shengyao, Xie Yaoni. Grounded theory: A New Qualitative Research Method [J]. Modern Education Series. 2001,(6): 51-53.
[6] Yang Yue. Research Methodology of Pharmaceutical Management [D] Shenyang: Shenyang Pharmaceutical University, 2004.
[7] Yang Yanyin. Research on drug re-registration and adverse drug reaction reporting and monitoring in Chongqing based on Grounded theory [D] Chongqing Medical University, 2013.
[8] Cai Mingyuan, Zhang Cuilian, Zuo Wei. Nationwide investigation and problems analysis of pharmaceutical preparations in medical institutions [J] China Pharmacy, 2022, 33(5):513-518.
[9] Zhu Yannian, Analysis of Measures to Strengthen Management and Standardize the Preparation of Medical Institutions under the Current Economic Situation [J]. Marketing Industry, 2021,(Z5):76-78.
[10] Zhao Ning, Shen Lin, Yu Zhen, et al. Analysis on the Usage and Prospect of Hospital Preparation Based on Literature [J] Guiding Journal of Traditional Chinese Medicine and Pharmacology,2021, 27(12): 59-64.
[11] Bao Changjun.Issues to be noted in the design and analysis of cross-sectional surveys [J] Jiangsu Preventive Medicine. 2007,18 (2):83-86.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230174

The Impact of Advertising Visibility on Consumers' Online Impulse Buying Behavior

Zijie Chen [1] , Rosita Binti Mohd Tajuddin, Jing Deng and Benhui Ren
College of Creative Arts, University Technology Mara (UiTM), 40450 Shah Alam, Malaysia

Abstract. With the popularization of online shopping, more and more advertisements are being placed on online shopping platforms. The impact of advertisements on consumers' online impulsive buying behavior has attracted much attention, and the phenomenon of consumers making online impulsive purchases is becoming more and more common. This field has become a hot research topic for scholars. This article aims to summarize the connotation of visible advertising, summarize indicators of consumer online impulse buying behavior, compare and analyze the impact of visible advertising on consumer online impulse buying behavior, and provide management decision-making reference for advertisers and e-commerce platforms.

Keywords. Online shopping, Advertising visibility, Impulsive buying behavior

1. Introduction

Impulsive buying is caused by a sudden desire to purchase, usually without a detailed purchase plan in advance, without considering long-term needs, and is an irrational behavior [1]. In the process of impulsive purchasing, consumers tend to focus their attention on the advantages of the product and overlook its drawbacks, resulting in the inability to objectively recognize and evaluate the product's attributes [2]. In impulse buying, there are many conflicting considerations. Impulsive buying refers to consumers who make purchases without a plan and cannot avoid it [3]. After being stimulated by the external environment, consumers ignite a strong and uncontrollable desire to purchase, which is a purchasing behavior that occurs as self-control weakens [4]. In addition, it may also be affected by advertising visibility, as advances in information technology and the rapid development of e-commerce have led to a shift from offline to online purchasing behavior. Transforming consumers to meet their daily needs through online shopping [5]. Therefore, this opens up opportunities for traders in the market to meet consumer shopping needs through online systems without having to go to the market location [5]. Every company must adjust its marketing strategy by entering online systems to sell its products [6]. Nowadays, online impulse buying is increasingly becoming a trend. It is estimated that approximately 40% of online consumer spending

[1] Corresponding author, Zijie Chen, College of Creative Arts, University Technology Mara (UiTM), 40450 Shah Alam, Malaysia; E-mail: huang87956673@163.com.

is caused by impulsive online purchases [7]. Online impulse buying is a sudden, immediate, and unplanned behavior of consumers online [8]. It can be concluded that online shopping is an increasingly growing trend today, especially impulse buying. Therefore, with the rapid competition between online stores today, ideal online marketers need to consider the factors that affect consumers' online shopping. One factor to consider is advertising visibility. This study suggests that the effectiveness of advertising and brand awareness has a direct positive and meaningful impact on purchase intention [9]. Social media marketing and advertising effects indirectly have a positive and significant impact on brand awareness purchasing intentions.

2. Analysis of current research status

2.1. Analysis of Research on Visible Advertising

Advertising visibility is an important component of measuring the effectiveness of display advertising. Advertisements on e-commerce platform web pages are not seen by consumers on the internet, and invisible or only partially visible advertisements do not have the opportunity to reach consumers, which will lead to ineffective advertising placement. Therefore, an increasing number of advertisements mainly seek to evaluate the value of displayed advertisements from the perspective of whether they are visible, and hope to pay only through visible exposure, rather than web page access requests for exposure. Advertisers are also beginning to try to measure whether the advertisements they display are visible.

The industry average reported by ComScore in 2013 was 54% of online advertising that is invisible, while Google's average by media is 50.2%. After intense discussions and experiments in mature foreign markets, in 2014, The Media Rating Council (MRC), in collaboration with the Interactive Advertising Bureau (IAB) of the United States, has jointly released the industry's advertising visibility measurement standards [10]: "For PC display advertisements, display at least 50% pixels in the visual area and display at least 1 second, especially for video advertisements, the display time is set to at least 2 seconds. The standards proposed by MRC have been understood and recognized by mainstream internet advertising platforms.

With the increasing attention paid to advertising in the industry, domestic and foreign scholars have also published some literature on technology. Zhang Weinan and others conducted empirical research on the measurement of visible exposure in display advertisements [11]. Usually, when an advertisement is loaded into a browser's webpage, it is considered an advertisement exposure. The article measures and determines the precise exposure times of displayed advertisements from two dimensions: the percentage of displayed pixels and the exposure time. It is concluded that more than 75% of displayed advertisements are displayed, and the display duration is at least 2 consecutive seconds, which can be considered as visible exposure. Wang Chong and others proposed collecting data from the media side to study the visual prediction of users' scrolling depth on the page, that is, there may be advertisements at the scrolling position of the page, which can predict the visibility of advertisements [12]. This article proposes two new models to predict the visibility probability under specific surface depths and verifies their advantages compared to other models. Bounie David and others conducted a comprehensive economic analysis of online advertising revenue after the use of visual technology. They discussed the impact of visual technology on advertising display, advertising prices, media revenue, and user experience [13]. Xie Xing and others introduced the process of image advertising platforms for image search and image upload. This platform can set different weights for images or image hotspots, and combine them with the platform's advertising library for advertising matching or bidding [14].

In addition, advertisers envision the practical use of visual display advertising: for example, visible exposure serves as the basis for purchasing in display advertising purchasing activities; Alternatively, by measuring the visibility of media advertising space, optimizing page layout design to balance advertising revenue and user experience; The data related to display advertising visibility is applied to the real-time advertising bidding process, which means that for exposure with high visibility, higher bids can be made [15]. However, ultimately, the importance of display advertising lies in its feasibility. Google refers to the standard of measuring whether displayed advertisements are actually viewed as "visible", increasing its visibility as the universal currency for advertising purchases, and making some progress in its business processes [16]. Since Google launched a visual based advertising purchasing service in its display advertising network, it has become an important factor for advertisers and advertising agencies to purchase advertisements.

At present, there is relatively little research on advertising visibility in the Chinese market. Only a few advertising platforms in the advertising ecosystem provide visibility measurement data, and due to differences in measurement methods, visibility data lacks transparency. The use of visible exposure as a delivery indicator has not yet formed a unified understanding, so it is necessary to introduce foreign theories and research results, in order to trigger further research and practical applications in different situations.

2.2. Analysis of Research on Impulsive Buying Behavior

Hawkins Stern's impulse buying theory, named after its proposer, Hawkins Stern, was proposed in 1962. This theory provides a different perspective on consumer purchasing behavior from most contemporary consumer behavior theories. Stern claims that customers' impulsive buying behavior is caused by external influences. According to this theory, marketers may persuade customers to purchase more products than they expected [17]. A recent survey targeting American customers found that 80% of respondents made impulsive purchases online, confirming previous findings [18]. Digital marketers are skillfully blending technology with their marketing goals to induce spontaneous purchases by the target audience [19]. Impulsive buying is a complex consumer behavior. In this process, consumers often make sudden decisions due to the attractiveness of external things. Due to strong sudden attraction, consumers often lack detailed and repetitive thinking in the decision-making process, often making quick and temporary decisions with strong impulsiveness. Cobb and Hoyer found that when consumers do not have a clear purchasing awareness, impulse buying is usually a temporary decision made by consumers [20]. Rook believes that impulsive buying behavior is caused by fluctuations in consumer emotions, which is usually a form of uncontrollable behavior. From the perspective of psychology, the impulsive purchasing behavior of consumers can be divided into two aspects: psychological impulse and impulsive behavior [21]. Wood found that impulse buying is mainly due to consumers lacking a certain degree of willpower. He believes that impulse buying generally includes two forms: free choice of purchasing behavior but lack of optimal judgment; Another type is compulsive impulse buying caused by compulsive will [22]. Therefore, this theory supports the impact of advertising visibility on customer impulse buying behavior.

The online impulse buying behavior of consumers may be caused by changes in external factors, thereby stimulating their emotions. Ning and others constructed a factor model for the impulse buying intention of online group buying consumers. Through questionnaire surveys and data analysis, it was concluded that consumer impulse buying has a positive impact on product characteristics, store environment, and consumer personality traits [23]. Li Yun studied the influencing factors of college students' impulse buying intention in the online purchasing environment, divided the environmental factors that promote college students' impulse buying into internal and external parts, and constructed corresponding theoretical models. The research results indicate that college students' personality traits, multi effect interactions, and diversity with the external world are more likely to encourage consumers to engage in impulsive consumption; The external factors of websites as shopping channels will not affect consumers' impulsive consumption [24]. Li Yifan conducted a study on impulsive consumption through online promotion methods. During the research process, we found that product factors, external environmental factors, individual characteristics of consumers, and consumption scenarios all affect consumers' impulsive purchasing behavior. Through a questionnaire survey of college students, it was found that promotional activities such as special offers, discounts, special offers, and lotteries have a positive impact on consumers' impulsive purchasing behavior [25]. Zhao and others explored the behavioral mechanism of consumers' online impulse buying. His research suggests that consumers' online impulse buying is influenced visually by website design, and product descriptions, website navigation, and visual display design of online stores can promote consumers' impulse buying [26]. Gan Xueping studied the impact of social media on consumers' impulse buying intention. Based on the influence of social media, it decomposes social media into six main influencing factors. Through empirical data research, social media has a positive impact on consumers' impulse buying intention through their perceived value. Research on consumer impulse buying behavior has shown that consumers' product characteristics, interests, attractiveness, interaction, and personal characteristics all affect their impulse buying intentions [27]. Gong and others studied the impact of online live streaming atmosphere on consumers' impulse consumption intention and found that atmosphere cues have a significant positive impact on impulse consumption intention; The atmosphere cues have a significant positive impact on the flow experience; The mobile experience has a positive impact on impulsive consumption intention[28].

3. Summary

Based on the above research analysis and summary, this article proposes certain marketing suggestions for e-commerce platforms and merchants, mainly in the following aspects:

Develop flexible and diverse online marketing strategies. Related studies have shown that product characteristics and store environment have a positive impact on consumers' impulsive purchase intention. Most consumers see advertisements and pages related to product features during online shopping, which stimulate their emotions and form impulsive purchasing behavior. Therefore, increasing the visibility of product features and store environment advertisements can attract more consumers' attention. On the other hand, in addition to increasing the visibility of regular discount advertisements during regular time periods, it is also possible to increase the visibility of event advertisements such as giving small gifts, sending coupons, and full discounts. In addition, centralized advertising promotion activities will be carried out in conjunction with specific festivals, such as the Double Eleven and Double Twelve Shopping Carnival, Spring Festival, Christmas, etc. On the other hand, combined with the popular online shopping live streaming as a means of communication, visual marketing stimulation is added to create a sense of integration and experience for consumers. During live streaming, advertising arrangements should be concise, paired with warm colors to create a sense of comfort and generate more positive emotions.

Increase the visibility of interactive advertising in online shopping. The latest research has found that when consumers perceive high interactivity, it can lead to higher impulse buying intentions. Firstly, the page design of online shopping platforms should be hierarchical and easy to operate. Display and display information such as product types and promotional activities, in order to enable consumers to quickly and conveniently find the products they need. Secondly, when consumers watch interactive advertisements, businesses should respond quickly and form an effective interactive environment. In some online shopping platforms with high traffic, interactive advertising has high visibility, and consumers can interact more with merchants. Sometimes, questions raised by merchants are answered in a timely manner, which can improve the shopping experience of consumers. At the same time, combining the enthusiastic attitude and sincere expression of the host in modern live streaming can enhance consumers' liking for them. Finally, interactive advertising should have the function of providing personalized services, paying attention to consumers' needs, and then providing appropriate products based on their needs, so that consumers feel sufficient respect and attention.

Segmenting the characteristics of consumers and implementing differentiation strategies. Based on relevant research, it has been found that when consumers have high impulsive traits, they are more likely to be motivated to develop impulsive purchasing intentions and further transform into impulsive purchasing behavior. Therefore, enterprises should enhance their attention recognition and understanding of these consumers' purchasing preferences and habits. Based on information acquisition and analysis from several aspects such as the duration of consumers watching short advertising videos, the transaction duration of purchasing goods, and the types of goods purchased, appropriate marketing strategies can be adopted for different types of consumers to maintain customer stickiness.

References

[1] Baumeister RF. Yielding to temptation: Self-control failure, impulsive purchasing, and consumer behavior. Journal of consumer Research. 2002;28(4):670–6.

[2] Li Yalin, Jing Fengjie. The Impact of Merchant Promotion Stimulation on Consumer Impulsive Post purchase Satisfaction: From the Perspective of Promotion Methods and Time Changes. Management Review. 2014;26(3):81–8.

[3] Solomon M, Russell-Bennett R, Previte J. Consumer behaviour: Buying, having, being. 2013;

[4] Liu Mengru. Research on the Influence of Website Featureson Impulse Buying [internet] [Master]. Nanjing University of Science and Technology; 2017 [Cited on May 15, 2023]. Contained in: https://kns.cnki.net/kcms2/article/abstract?v=3uoqIhG8C475KOm_zrgu4lQARvep2SAkVtq-vp-8QbjqyhlE-4l1YmMZLXZ46w0Lb1wISSbyVTP3YlW-G0CoQfI6Be_YS6Yf&uniplatform=NZKPT

[5] Harahap DA. Pembatasan sosial berskala besar (psbb) menangani pandemi covid-19 dan tren pembelian online. 2020;

[6] Harahap DA. Perilaku belanja online di Indonesia: Studi kasus. JRMSI-Jurnal Riset Manajemen Sains Indonesia. 2018;9(2):193–213.

[7] Liu Y, Li H, Hu F. Website attributes in urging online impulse purchase: An empirical investigation on consumer perceptions. Decision support systems. 2013;55(3):829–37.
[8] Piron F. Defining impulse purchasing. ACR North American Advances. 1991;
[9] Maria S, Pusriadi T, Hakim YP, Darma DC. The effect of social media marketing, word of mouth, and effectiveness of advertising on brand awareness and intention to buy. Jurnal Manajemen Indonesia. 2019;19(2):107–22.
[10] Wang C, Kalra A, Borcea C, Chen Y. Viewability prediction for online display ads. Included in: Proceedings of the 24th acm international on conference on information and knowledge management. 2015;413–22.
[11] Zhang W, Pan Y, Zhou T, Wang J. An empirical study on display ad impression viewability measurements. arXiv preprint arXiv:150505788. 2015;
[12] Wang C, Kalra A, Zhou L, Borcea C, Chen Y. Probabilistic models for ad viewability prediction on the web. IEEE Transactions on Knowledge and Data Engineering. 2017;29(9):2012–25.
[13] Bounie D, Valérie M, Quinn M. Do you see what I see? Ad viewability and the economics of online advertising. Ad Viewability and the Economics of Online Advertising (March 1, 2017). 2017;
[14] Xie X, Scott MR, Liu D, Jiang W, Tien J. Image-Based Advertisement Platform. Google Patents; 2010.
[15] Tautenhahn R, Böttcher C, Neumann S. Highly sensitive feature detection for high resolution LC/MS. BMC bioinformatics. 2008;9(1):1–16.
[16] Mohan N. How to make advertising visibility truly visible? China Advertising. 2015;(6):150.
[17] Dutta T, Mandal MK. Neuromarketing in India: understanding the indian consumer. Routledge; 2018.
[18] Verhagen T, Van Dolen W. The influence of online store beliefs on consumer online impulse buying: A model and empirical application. Information & Management. 2011;48(8):320–7.
[19] Carter S. 80% of younger shoppers make impulse purchases online-here's show sites trick you into spending. 2018.
[20] Cobb CJ, Hoyer WD. Planned versus impulse purchase behavior. Journal of retailing. 1986;
[21] Rook DW. The buying impulse. Journal of consumer research. 1987;14(2):189–99.
[22] Wood M. Socio-economic status, delay of gratification, and impulse buying. Journal of economic psychology. 1998;19(3):295–320.
[23] Ning Lianju, Zhang Xinxin. An Empirical Study on the Factors Influencing Consumers' Impulsive Purchase Intention in Online Group Buying. Journal of Fujian Normal University(Philosophy and Social Sciences Edition). 2011;(6):13–8.
[24] Li Yun. Analysis on the Influencing Factors of College Students' Impulse Shopping [internet] [Master]. Chongqing University; 2016 [Cited on May 15, 2023]. Contained in: https://kns.cnki.net/kcms2/article/abstract?v=3uoqIhG8C475KOm_zrgu4lQARvep2SAkkyu7xrzFWukWIylgpWWcEkGItWT2Pi_VtLPRnRV4GoJ4e9Fm8Xfl97umDu9D3Z1M&uniplatform=NZKPT
[25] Li Yifan. The Impact of Online Promotion Methods on Impulsive Consumption Behavior of College Students: A Case Study of Universities in Guangzhou University City. Modern Marketing (Business Edition). 2019;(9):76–7.
[26] Zhao Yong, Tan Gang, Yang Yanlin. Analysis and Management of the Psychological Mechanism of Consumer Impulsive Purchase Behavior. Technology Entrepreneurship Monthly. 2008;(1):2.
[27] Gan Xueping. A research on the effect of social media impulsive on consumer's buying intention [internet] [Master]. Hubei University of Technology; 2013 [Cited on May 15, 2023]. Contained in: https://kns.cnki.net/kcms2/article/abstract?v=3uoqIhG8C475KOm_zrgu4lQARvep2SAk9z9MrcM-rOU4mSkGl_LWf7-L-yiUWKiIyggjUzT8p9qPPNeHGZDtsGql7aBeL5UU&uniplatform=NZKPT
[28] Gong Xiaoxiao, Ye Zuoliang, Wu Yuping, Liu Jiaying. A Study on the Mechanism of the Influence of Live Scene Atmosphere Clues on Consumer Impulsive Consumption Intention. Journal of Management. 2019;16(6):875–82.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230175

Research on Industrial Internet Talent Training Based on Enterprise Application Demand

Nana Shan[a,1] and Yi Qian[a], Xiujuan Sun[a]
[a] *Affiliation School of Information Science and Technology, Taishan University, China*

Abstract. The Industrial Internet can be used in enterprises for predictive maintenance, location tracking, workplace analysis, remote quality monitoring and energy optimization. To realize the development of the industrial Internet, a large number of high-quality industrial Internet talents are indispensable. These talents not only need to have a solid foundation of engineering technology and a good level of computer technology but also need to have a deep understanding and grasp of the industrial production process and be able to develop suitable industrial Internet solutions according to actual needs, to promote the steady development of the industrial Internet. Based on preliminary research and communication with enterprises, this paper designed some courses for Industrial Internet Talent Training.

Keywords. Industrial Internet, Talent Training, Enterprise Application Demand

1. Introduction

As an important part of the new generation of the information technology industry, the industrial Internet is not only an important part of the new infrastructure but also the main way and means of digital transformation of the real economy. The innovation and development of the industrial Internet have become a national strategy. Industrial Internet is a new type of infrastructure, application mode, and industrial ecology formed by the all-round and deep integration of the new generation of information technology and industrial system, which mainly consists of four parts: network, platform, security, and data. Among them, the network is the foundation, data is the element, the platform is the core, and security is the guarantee.

Industrial Internet focuses on key industries such as equipment, electronics, chemical industry, mining, food, pharmaceuticals, steel, energy, construction, and transportation, providing all-cloud and all-scene solutions for the intelligent transformation of the manufacturing industry, breaking through core technologies such as identification analysis, industrial safety, and deterministic network.

However, based on the rapid development of the industrial Internet and related industries and technologies, the existing curriculum content and teaching methods can

---

[1] Nana Shan; School of Information Science and Technology, Taishan University. E-mail: helensnn@hotmail.com. This work was supported by teaching reform research of Taishan University (JG202116) and Special project of education science planning in Tai'an Classroom teaching based on teaching reflection (TJK2202106ZX039).

no longer meet the needs, and the talent training mode urgently needs to be connected with industries and industries [1].

To fully meet the demand of industrial Internet talent, we should innovate and carry out the cooperation of "industry, education, and research", carry out the strategy of strengthening the country by talents, promote the construction of technical and skilled personnel in the new vocational field, and promote the training of talents in the new economic field. Higher education personnel training also needs to fully connect with industry and enterprise research, and carry out the reform of curriculum content and teaching methods [2].

2. Research status and current problems

According to strategic positioning and planning, the industrial Internet includes five major businesses [3]. As Table 1 shows:

Table 1. five major businesses

No.	Name	Discribe
1	industrial equipment	including special industrial equipment, digital equipment after digital transformation, and intelligent products upgraded through active labeling, to become industrial equipment layer product service providers
2	industrial communication	including identification analysis, spark chain network facilities, enterprise internal and external network hybrid networking and 5G + industrial Internet, including industrial deterministic network
3	industrial Internet platform	which mainly solves the cross-industry and cross-domain problems of enterprises, and creates many platforms with industry and regional characteristics and professional platforms
4	industrial software	including industrial design software, business management software, industrial control software, and embedded software
5	industrial safety	By creating industrial security products, it provides five layers of security to ensure the data security of enterprises from R & D, design to production and sales.

The key to the transformation and upgrading of the industrial economy is talent, which is the core element of the development of the industrial Internet. However, there are many problems in the construction of the industrial Internet talent team, which has become the main "bottleneck" restricting the rapid development of the industrial Internet. The industrial Internet is the deep integration of the new generation of information technology and the manufacturing industry, and the landing application requires cross-border talents who understand both industrial operation and network information technology. The industrial Internet education should be theoretical education, which requires hundreds of years of industrial experience to be transformed into useful knowledge to reuse and software, and its technological innovation and application need to be coordinated in many fields and disciplines, including software engineering, enterprise management, product design, robotics, and equipment [4].

3. Research course content and teaching method

This paper will study the training of industrial Internet talents from the following two aspects:

3.1. Research on course content docking with industry enterprises

In terms of course content, according to the software development needs of enterprises such as industrial control and equipment management, new engineering courses such as rapid software development and other basic, cross-cutting, and professional courses can be offered [5].

Based on preliminary research and communication with enterprises [6], the following rapid development courses are listed:

Table 2. the list of courses

No.	Course name	Content	Class hours
1	An overview of container technology	PPT/Video	2
2	Container product introduction	PPT/Video	2
3	Fundamentals of DevOps	PPT/Video	2
4	DevOps Datasheet	PPT/Video	2
5	DevOps Field Camp	PPT/Video	2
6	Agile development management	PPT/Video	2
7	Continuous delivery	PPT/Video	2
8	Technical operation	PPT/Video	2
9	Microservices	PPT/Video	2
10	Service Grid	PPT/Video	2
11	Micro-service practice	PPT/Video	4
12	Big Data Foundation	PPT/Video	4
13	Hive Data Warehouse Tool Technology	PPT/Video	4
14	HBase Real-time Distributed NoSQl Database	PPT/Video	4
15	Spark Big Data Processing Technology and Application	PPT/Video	4
16	Practical Cases of Industrial Internet Application in Different Industries and Fields	PPT/Video	4

3.2. Research on teaching methods docking with industry and enterprises

In terms of teaching methods, in addition to traditional classroom teaching and practical teaching, we can also use the industrial Internet platform of enterprises to complete developer tasks and carry out new teaching methods reform [7]. Relying on the dominant disciplines, carry out the development and research of practical training and practice in related fields, including but not limited to the industrial Internet equipment management or control software involved in the current industrial Internet platform [8]. So, we can make full use of the platform of enterprises and help them to improve their product.

Table 3. Introduction to the courses

No.	categories	name	Introduction to the course
1	Industrial Internet of Things Training Platform	Industrial Internet of Things technology	Provide IOT equipment demonstration, in-depth understanding of IoT and hardware, wireless sensor network application technology, RFID application technology and other basic knowledge and skills, so that students can master the basic knowledge and application scenarios of IoT, and learn how to build industrial IoT based on typical manufacturing production lines [9].
2		Industrial Internet of Things Platform	Demonstrate the industrial Internet of Things platform, understand the platform architecture, understand the industrial data acquisition and data analysis module, and operate the data acquisition and data analysis process. To enable students to master the principles and methods of industrial data acquisition, data processing and presentation. Functional scenarios such as real-time monitoring of factory environment and monitoring of electricity consumption are simulated [10].
3	Industrial Application Development Training Platform	Industrial Application Development Training Platform	The training platform provides a series of basic tools and components for application and integration. Quickly generate and integrate application software based on the development mode of no code or a very small amount of code on the platform. To enable students to master the process of development and integration of industrial applications [11].
4	Industrial Big Data Training Platform	Build industry data lake/data center	The IBP data factory supports data access of multiple data sources and multiple data types, and builds a data lake through efficient processing, which is convenient for users to manage data in a unified and convenient way. At the same time, IBP data factory has accumulated rich industry models and algorithms in various business areas to analyze the data in the lake. Through practical training, students can master the data components, data source configuration and data cleaning functions of IBP data factory, and cultivate the ability to build a data platform [12].
5		Massive data computing	Capacity expansion and shrinkage of the cluster, providing a variety of relational databases and big data databases, and providing big data computing components such as MapReduce, Spark and SparkSQL. Through practical training, students can choose the appropriate system architecture to deploy the system for the customer according to the actual business scenario of the customer, and use the appropriate big data computing components to meet the customer's requirements for timeliness of business data computing [13].
6		Real-time streaming data analysis	Analyze monitoring real-time stream data and support online monitoring alarm; The analysis of traffic flow data supports real-time display of traffic hot spots, optimization of signal timing and guidance of driving routes. The analysis of user usage data supports real-time recommendation and optimizes the recommendation algorithm.

7		Data acquisition exchange	The industrial big data training platform supports the access of multiple database data sources, including but not limited to Oracle, MySQL, SQLServer, DB2, Hangao and other mainstream relational databases at home and abroad, HBase, MongoDB and other NoSQL data sources; At the same time, it supports a variety of different types of structured, semi-structured and unstructured data access. Support business system data collection, reporting and exchange requirements. Through practical training, students can master the technology of big data aggregation [14].
8		Visual application developme nt	Industrial big data training platform has built-in rich data visualization components, while meeting the needs of business charts, dashboards, reports and large screens, supporting the construction of business application visualization. Through practical training, students can master rich data visualization components, and have business application visualization construction according to the needs of business charts, dashboards, reports and large screens.
9	Intelligent Manufactur ing Training Platform	Practical Course of Introductio n to Intelligent Manufactur ing Engineerin g	System analysis method, system diagnosis method and system implementation for learning intelligent manufacturing engineering
10		Fundament als of Intelligent Manufactur ing Production Manageme nt	Understand the production management concepts such as assembly line, process and yield; Understand the PDCA process in the production process; Be able to make production plans for small production lines and realize visual display; Understand the concept of line balancing and be able to identify bottleneck processes; Conscious of quality, on-time delivery and cost control; Able to read and fill in production reports.

4. The Discussion and Conclusion

With the rapid development of a new round of scientific and technological revolution and industrial transformation, the Internet has extended rapidly from the consumption field to the production field, the industrial economy has expanded from digitalization to networking and intellectualization, and the innovative development of the Internet and the new industrial revolution have formed a historic intersection, giving birth to the industrial Internet. Accelerating the development of the industrial Internet and promoting the deep integration of the new generation of information technology and the manufacturing industry is the key to conforming to the trend of technological and industrial change, speeding up the construction of manufacturing power and network power, to deepen the structural reform of supply side, to promote the transformation and upgrading of the real economy, and to achieve the goal of "carbon peak, carbon neutralization" Therefore, with the continuous development of the industrial Internet technology, the demand for industrial Internet talents will be growing in the future. It is predicted that by 2025, the talent gap in the core industry of the industrial Internet will reach about 2.54 million. It is more and more important to strengthen the training of industrial Internet talents and continuously improve their skills and comprehensive quality. Through the cooperation and research with many enterprises and universities in the early stage, this paper proposed the reform plan of the curriculum system and teaching method of industrial Internet talent training, hoping to meet the rapidly growing needs of industrial Internet talent.

References

[1] Khan, A.R., Chishti, M.A. βDSC2DAM: beta-dominating set centered Cluster-Based Data Aggregation mechanism for the Internet of Things. J Ambient Intell Human Comput 13, 4279–4296 (2022).
[2] Yang, Y. , Zheng, D. , Shi, W. , Gong, M. , & University, Y. . (2019). Research on innovative teaching reform of c language programming for training of new engineering talent. Computer Era.
[3] China Academy of Information and Communications Technology. Industrial Internet industry economic development report. 2020.
[4] CISCO (2020) Cisco annual internet report (2018–2023) White paper. https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html. Accessed 24 Feb 2020.
[5] Wang L , Gao T , Zhou B , et al. Manufacturing service recommendation method toward industrial internet platform considering the cooperative relationship among enterprises. Expert Systems with Applications, 2022, 192:.
[6] Dm A , Np B , Ja B . Production management guided by industrial internet of things and adaptive scheduling in smart factories. Design and Operation of Production Networks for Mass Personalization in the Era of Cloud Technology, 2022:117-152.
[7] Chai T Y , Liu Q , Ding J L , et al. Perspectives on industrial-internet-driven intelligent optimizedmanufacturing mode for process industries. SCIENTIA SINICA Technologica, 2022, 52(1):14-25.
[8] Huo R , Zeng S , Wang Z , et al. A Comprehensive Survey on Blockchain in Industrial Internet of Things: Motivations, Research Progresses, and Future Challenges. Communications surveys & tutorials, 2022(1):24.
[9] Sun J , Yang T , Xu Z . Assessing the implementation feasibility of intelligent production systems based on cloud computing, industrial internet of things and business social networks. Kybernetes: The International Journal of Systems & Cybernetics, 2022(6):51.
[10] Wang T , Ai S , Tian Z , et al. A Blockchain-Based Distributed Computational Resource Trading System for Industrial Internet of Things Considering Multiple Preferences. arXiv e-prints, 2022.
[11] Tariq U , Ahanger T A , Ibrahim A , et al. The Industrial Internet of Things (IIoT): An Anomaly Identification and Countermeasure Method[J]. Journal of Circuits, Systems and Computers, 2022, 31(12).
[12] LiuChao, SuZiwei, XuXun, et al. Service-oriented industrial internet of things gateway for cloud manufacturing. 2022.
[13] Armin, Balalaie, Abbas, et al. Microservices Architecture Enables DevOps: Migration to a Cloud-Native Architecture[J]. IEEE Software, 2016, 33(3):42-52.
[14] Haindl P , Plsch R . Focus Areas, Themes, and Objectives of Non-Functional Requirements in DevOps: A Systematic Mapping Study. 2022.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230176

Research on the Influence of Tax Incentives on the Innovation Performance of Pharmaceutical Manufacturing Enterprises

Xia Ouyang[a], Can Li[a] and Dan Li[a,1]
[a] *School of Economics and Management, Xiangnan University, Chenzhou, China*

Abstract. Taking 28 listed companies in China's A-share pharmaceutical manufacturing industry from 2018 to 2021 as samples, with tax incentives as an explanatory variable, R&D investment intensity as an intervening variable, enterprise size, enterprise establishment time, return on total assets, debt-to-assets ratio, and fixed asset density as control variables, and innovation performance as explained variable, a multiple linear regression model is established, and descriptive statistics, correlation analysis, multicollinearity test, and robustness test are carried out to empirically analyze the influence of tax incentives on the innovation performance of pharmaceutical manufacturing enterprises. The results show that tax incentives have a significant positive impact on the innovation performance of pharmaceutical manufacturing enterprises, and tax incentives can promote pharmaceutical manufacturing enterprises to increase the intensity of R&D investment, and the intensity of R&D investment has a certain mediating effect and long-term influence on the relationship 1between tax incentives and innovation performance. Combined with the empirical results, the paper puts forward the strategy of improving enterprise innovation performance to enhance the core competitiveness of enterprises.

Keywords. Pharmaceutical Enterprises, Tax Incentives, R&D Investment Intensity, Innovation Performance

1.Introduction

Innovation is the internal driving force of enterprise development, the level of innovation performance is related to the sustainable development of enterprises. Tax incentives are the main means for the government to guide enterprise innovation. The government grants tax relief to specific taxpayers, taxpayers or taxpaying regions through tax means, which can effectively reduce the high tax burden of enterprises, and preferential tax policies allocate innovative resources through the market. Therefore, China has introduced a series of preferential tax policies such as income tax reductions

---

[1] Corresponding Author, Dan Li, School of Economics and Management, Xiangnan University, Chenzhou, China; E-mail:187237129@qq.com.
It is the research result of The Key Cultivation Base for “The 14th Five-Year Plan” of Educational and Scientific Research (Lifelong Education Research Base (Fundamental Theory Area)) (grant no. XJK22ZDJD58) in Hunan Province.

and exemptions and super deduction of research and development expenses to support and encourage enterprises to innovate and develop.

domestic and foreign scholars have carried out extensive research on the impact of tax incentives on corporate innovation performance, and most scholars have verified the incentive effect of tax incentives on innovation performance through empirical research. For example, Wang and Schmidt (2002) added the dummy variable of the panel unit to the Tobit model, and concluded that the tax policy can have a positive incentive effect on the innovation output performance of technology-based enterprises [1]. Michael Peneder (2008) thought that the preferential tax policy expands the coverage, mobilizes more enterprises' enthusiasm for R&D and innovation, and then positively spills over to the whole society [2]. Hongxiang Tang and Yinchang Li (2020) conducted research on China's GEM industrial enterprises from 2015 to 2019 as sample companies, and found that there is a significant positive correlation between tax incentives and corporate innovation performance, and tax incentives can effectively improve corporate innovation performance [3]. Hui Xing et al. (2021) systematically conducted empirical research using methods such as DID and PSM, and the results showed that preferential tax policies promote substantial innovation of enterprises through three functions [4].

In addition, research on the ways in which tax incentives affect corporate innovation performance has also attracted much attention. Scholars believe that tax incentives can promote the R&D investment of enterprises, and the increase of R&D investment can stimulate the innovation motivation of enterprises to improve innovation performance, R&D investment can play a mediating effect in the relationship between tax incentives and innovation performance. Fabiani et al. (2014) found that tax incentives stimulate enterprises' R&D investment by reducing innovation costs through their research on private enterprises in Brazil [5]. Andrew C (2018) conducted an empirical study on the financial data of American technology companies and found that for every percentage point increase in the tax incentives enjoyed by companies engaged in technology research and development, their main R&D investment will increase by 2.8%-3.8%[6]. Jingyi Li et al. (2020) studied on the high-tech enterprises in the western region and found that the interaction between tax incentives and enterprise R&D investment can stimulate the innovation performance of enterprises [7].Congrong Li et al. (2021) pointed out that tax incentives have an effect on innovation performance by affecting R&D investment in their research on the impact of fiscal policy on the innovation performance of military-civilian integration enterprises [8].

Scholars mostly choose enterprises in high-tech industries as their research objects, and there are relatively few studies on the impact of tax incentives on innovation performance of enterprises in other industries. Based on this, this paper takes 28 listed companies in China's pharmaceutical manufacturing industry as samples, puts forward research hypotheses on the basis of domestic and foreign research results, and constructs a multiple regression model to empirically analyze the relationship between tax incentives, R&D investment and innovation performance of pharmaceutical manufacturing companies, so as to explore The impact mechanism and effect of tax incentives on the innovation performance of pharmaceutical manufacturing enterprises.

2.Research Design

2.1 Research Hypothesis

Because of its fair and transparent characteristics, preferential tax policies have become an important means to promote enterprise innovation, and they are widely used to encourage enterprise innovation. The implementation of tax incentives can greatly promote the innovation output and innovation efficiency of enterprises[9]. At the same time, tax incentives can subsidize the R&D expenses of enterprises through preferential tax rates and direct deductions, so as to achieve the purpose of improving the innovation performance of enterprises[10]. Therefore, this paper makes the following hypothesis :

Hypothesis 1: Tax incentives are positively correlated with enterprise innovation performance.

On the one hand, tax incentives can reduce the tax burden of beneficiary enterprises, which is equivalent to obtaining a sum of funds from the government for reinvestment. On the other hand, post-incentives such as super deduction of R&D expenses require enterprises to spend on R&D and innovation in the early stage, which can mobilize the enthusiasm of enterprises for R&D and increase the intensity of R&D investment. Junjiao Liang, Yuxi Jia (2019), Fenfen Miao(2021), etc. all believe that the promotion effect of tax cuts on R&D investment is more obvious [11][12]. Therefore, this paper proposes the following hypothesis :

Hypothesis 2: Tax incentives are positively correlated with R&D investment intensity.

According to the theory of technological innovation, enterprises can promote the smooth development of innovation activities by increasing R&D investment, so as to produce products with core competitiveness or improve their core technology level, thereby improving the innovation performance of enterprises. Yuchen Xiong and Yinguo Li (2019) 's research on Chinese technology companies shows that there is a significant correlation between their R&D investment and corporate innovation performance [13]. Dianchun Jiang and xiaowang Pan (2022) found that increasing R&D investment can bring more innovation output to enterprises [14]. Therefore, this paper proposes the following hypothesis :

Hypothesis 3:There is a positive correlation between innovation performance and R&D investment intensity.

Since policies are generally lagging, and the cycle of enterprise innovation is generally taking several years, this paper will deeply explore the impact of tax incentives and R&D investment on the innovation performance of pharmaceutical manufacturing companies under lagging conditions. Therefore, this paper proposes the following hypothesis :

Hypothesis 4: There is a time lag in the effect of R&D investment intensity on enterprise innovation performance.

2.2 Source of Sample Data

This paper selects the research samples of Shanghai and Shenzhen A-share listed companies in the pharmaceutical manufacturing industry, which is classified by China Securities Regulatory Commission. Referring to variable selection methods such as Qiulai Zhang(2018) and Mengyun Han(2020)[15][16],the number of invention patent applications, R&D expenses, business income, income tax expenses, total profit and other relevant raw data of sample companies from 2018 to 2021 were collected from CSMAR database, the official website of QiChaCha and Patentstar. Based on the integrity and accessibility of corporate financial information, data of 28 sample companies were obtained. EXCEL software was used to process the selected data, and then STATA software was used for empirical analysis.

2.3 Research Variables

a) Explained Variable

Innovation performance (PATENT) is measured as the natural logarithm of the number of invention patent applications. This index is used to measure an enterprise's ability to enhance the value of innovation results through technical solutions or methods and evaluate its innovation performance. The innovation performance of enterprises is growing as the number of invention patent applications is bigger.

b) Explanatory Variable

Tax incentives (TAX) are measured by the corporate income tax burden. This index is the ratio of the current income tax expense to the total income of the enterprise, which is negatively correlated with the level of tax incentives enjoyed by enterprises, that is, the smaller the corporate income tax burden is, the greater the level of tax incentives enjoyed by enterprises in the current year.

c) Intervening Variable

Research and development investment intensity (R&D) is measured by the ratio of R&D expenses to operating revenue. Enterprises support innovation activities through R&D expenditure. This index can eliminate the difference in the level of R&D investment caused by the different sizes of enterprises. The higher the ratio, the more importance enterprises attach to R&D.

d) Control Variables

Enterprise size (SIZE) is represented by the natural logarithm of the enterprise’s total assets at the end of the current period. It is generally believed that large enterprises have more innovative capabilities and resources and are relatively easy to obtain government innovation support.

The establishment time of the enterprise (YEAR) is represented by the statistical year minus the listing year of the enterprise plus one. The longer an enterprise is established, the more experience it has accumulated in carrying out innovation activities, which may influence the innovation research and development activities of the enterprise to a certain extent.

Return on total assets (ROA) is the ratio of net income after tax to total assets at the end of the period. The higher the ROA is, the stronger the enterprise is in terms of capital saving and income increase, and the better the profitability of the enterprise is, which is conducive to the enterprise investing more funds in R&D and innovation.

Debt-to-assets ratio (DAR) is the ratio of total ending liabilities to total ending assets. The smaller the DAR is, the easier it is for an enterprise to get external financial support, thus increasing R&D investment and accelerating the output of innovation achievements.

Fixed Asset Density (FAD) is the ratio of net fixed assets to total assets at the end of the period. The higher the FAD, the stronger the debt guarantee ability of enterprises, and the much more choices of innovative investment financing methods.

The specific research variables are introduced in Table 1:

Table 1. Introduction of Research Variables

Variable type	Variable identifier	Variable name	Calculating formula
Explained variable	PATENT	Number of invention patent applications	The natural logarithm of the number of invention patent applications
Explanatory variable	TAX	Corporate income tax burden	current income tax expense/total income
Intervening variable	R&D	Research and development investment intensity	R&D expenses/operating revenue
	SIZE	Enterprise size	The natural logarithm of the total assets of the enterprise at the end of the current period
Control variables	YEAR	Enterprise establishment time	Statistical year - the listing year of the enterprise + 1
	ROA	Return on total assets	Net income after tax/total ending assets
	DAR	Debt-to-assets ratio	Total ending liabilities/total ending assets
	FAD	Fixed Asset Density	Net fixed assets/total ending assets

2.4 Model Design

The following four multiple regression models are built for analysis according to the above assumptions.

To verify hypothesis 1: Tax incentives are positively correlated with enterprise innovation performance, Model 1 is constructed as Formula (1):

$$PATENTt_{i,t}=\alpha_0+\alpha_1*TAX_{i,t}+\Sigma\,\alpha_j*Control_{i,t}+\varepsilon_{i,t} \quad (1)$$

To verify hypothesis 2: Tax incentives are positively correlated with R&D investment intensity, Model 2 is constructed as Formula (2):

$$R\&D_{i,t}=\alpha_0+\alpha_1*TAX_{i,t}+\Sigma\,\alpha_j*Control_{i,t}+\varepsilon_{i,t} \quad (2)$$

To verify hypothesis 3: There is a positive correlation between innovation performance and R&D investment intensity, Model 3 is constructed as Formula (3):

$$PPATENT_{i,t}=\alpha_0+\alpha_1*R\&D_{i,t}+\Sigma\,\alpha_j*Control_{i,t}+\varepsilon_{i,t} \quad (3)$$

To verify hypothesis 4: There is a time lag in the effect of R&D investment intensity on enterprise innovation performance, Model 4 is constructed as Formula (4):

$$PATENT_{i,t}=\alpha_0+\alpha_1*R\&D_{i,t-n}+\Sigma\,\alpha_j*Control_{i,t}+\varepsilon_{i,t} \quad (4)$$

Control represents all control variables; i represents the i-th listed enterprise in the pharmaceutical manufacturing industry ($i\in[1, 28]$); t represents year ($t\in[2018, 2021]$); n represents the number of lag periods ($n\in[1, 2]$); $\varepsilon_{i,t}$ is the random error terms for the individual enterprise and time mix differences.

3.Empirical Test and Analysis

3.1 Descriptive Statistical Analysis

As can be seen in Table 2:

(1) The number of invention patent applications (PATENT). The minimum value, maximum value and average value of the number of invention patent applications of Chinese pharmaceutical manufacturing listed enterprises are 0.693, 5.056 and 2.505 after logarithm, indicating a low overall level. Its standard deviation is 1.002, indicating a large gap in innovation performance among sample enterprises in the pharmaceutical manufacturing industry.

(2) Corporate income tax burden (TAX). The minimum value, maximum value and standard deviation of the corporate income tax burden of sample enterprises are -0.00397, 0.237, and 0.0356, indicating data with high dispersion and negative value. The average tax burden of enterprise income tax is 0.111, indicating that the income tax burden of most enterprises is still within a reasonable range and enjoys a relatively high level of tax incentives, while only a few enterprises enjoy a significant difference in the level of tax incentives.

(3) Research and development investment intensity (R&D). The standard deviation of R&D investment intensity is 0.0379, the minimum value is 1%, the maximum value is 22.9%, and the mean is 3.8%, indicating that the overall competitiveness of Chinese pharmaceutical manufacturing enterprises is strong, but there are large differences in the R&D investment intensity among sample enterprises.

(4) Enterprise size (SIZE). Enterprise size's minimum value is 20.64, the maximum value 24.39, and the mean 22.54, indicating a small gap.

(5) Enterprise establishment time (YEAR). The mean value, standard deviation, minimum value and maximum value of the establishment time of the pharmaceutical manufacturing sample enterprises are 15.18, 5.811, 5.167 and 29.18, indicating a large gap in the development stage of the pharmaceutical manufacturing sample enterprises.

(6) Return on total assets (ROA). The average value of return on total assets is 0.0962, and the maximum value is 47 times of the minimum value, showing the net income after tax per unit of assets created by different enterprises varies greatly.

(7) Debt-to-assets ratio (DAR). The standard deviation of ADR is 0.117, the minimum value is 0.0420, the maximum value is 0.482, and the mean is 0.270. It is generally believed that the reasonable DAR is no more than 50%, so it can be seen that the gap in debt level of all sample pharmaceutical manufacturing enterprises is relatively small and at a low-risk level.

(8) Fixed Asset Density (FAD). The difference between the mean, minimum and maximum of the FAD is small, and the standard deviation is 0.0926, indicating that the gap between most sample enterprises is small.

Table 2. Statistical Analysis Results of Variable Description

Variable name	Observed value	Mean	Standard deviation	Minimum value	Maximum value
PATENT	112	2.505	1.002	0.693	5.056
TAX	112	0.139	0.036	-0.004	0.237
R&D	112	0.061	0.038	0.010	0.229
SIZE	112	22.540	0.827	20.640	24.390
YEAR	112	15.180	5.811	5.167	29.180
ROA	112	0.096	0.057	0.007	0.340
DAR	112	0.270	0.117	0.042	0.482
FAD	112	0.216	0.093	0.057	0.521

3.2 Correlation Analysis

Before multiple regression analysis, correlation analysis can describe the degree of correlation between each explanatory variable and the explained variable, as well as the linear correlation between each explanatory variable.

Table 3. Correlation Analysis Results of Variables

	PATENT	TAX	R&D	SIZE	YEAR	ROA	DAR	FAD
PATENT	1.000							
TAX	-0.160***	1.000						
R&D	0.388***	-0.358***	1.000					
SIZE	0.472***	-0.057	0.362***	1.000				
YEAR	0.165***	-0.071	0.114*	0.470***	1.000			
ROA	0.160***	-0.016	0.027	0.358***	0.209***	1.000		
DAR	-0.192***	-0.026	-0.383***	0.052	0.070	-0.217***	1.000	
FAD	-0.218***	-0.175***	-0.130**	-0.282***	-0.038	-0.312***	0.188***	1.000

***, **, * means the correlation is significant in 1%, 5%, 10% level respectively

As shown in Table 3, corporate income Tax burden (Tax), R&D investment intensity (R&D), enterprise size (Size), establishment time (Year), return on total assets (ROA), debt-asset ratio (DAR), fixed asset density (FAD) are all significantly correlated with Patent applications at the level of 1%. Among them, the tax burden of corporate income tax (Tax) is significantly negatively correlated with the number of patent applications (Patent) and the intensity of R&D investment. The number of patent applications is significantly positively correlated with the intensity of R&D investment.. It shows that tax incentives can stimulate pharmaceutical manufacturing enterprises to increase R&D investment, and R&D investment and tax incentives can promote the improvement of innovation performance of them, which preliminarily verifies hypotheses 1, 2 and 3.

3.3 Multicollinearity Test

In order to avoid multicollinearity, tolerance and variance inflation factor (VIF) analysis were used, and the results were shown in Table 4. Combined with Table 3 and Table 4, it can be seen that the absolute values of all variable coefficients are between 0 and 0.48, and the VIF between non-explained variables are all less than 2, indicating that there is no serious problem with multicollinearity between explanatory variables and control variables, and empirical analysis can be conducted.

Table 4. Results of Multicollinearity Test

Variable name	VIF	Tolerance
SIZE	1.92	0.519924
YEAR	1.49	0.669300
ROA	1.40	0.712960
DAR	1.32	0.759977
FAD	1.29	0.774239
SIZE	1.25	0.797877

3.4 Regression Analysis

Multiple regression analysis is conducted on the four models constructed in this paper, as shown in Table 5. All four models constructed in this paper pass the F test, indicating that the four regression models are significant. Among them, Model 1 and Model 2 are significant at the 1% level, and Model 3 and Model 4 (lag period 1 and lag period 2, respectively) are significant at the 5% level

Table 5. Model Regression Analysis Results

Explained variable / Explanatory variable	PATENT	R&D	PATENT (current)	PATENT (lag period 1)	PATENT (lag period 2)
TAX	-4.254*** (-1.40)	-0.385*** (-3.37)			
R&D			5.190** (1.92)	6.081** (1.57)	6.811** (1.24)
SIZE	0.694*** (5.10)	0.021*** (5.52)	0.593*** (3.87)	0.580*** (3.24)	0.560** (1.24)
YEAR	-0.010 (-0.53)	-0.000 (-0.66)	-0.008 (-0.43)	0.002 (0.06)	-0.010 (-0.32)
ROA	-1.754 (-1.20)	-0.168*** (-3.68)	-0.822 (-0.52)	-0.430 (-0.22)	1.052 (0.41)
DAR	-2.135*** (-2.64)	-0.148*** (-5.97)	-1.376 (-1.54)	-1.718 (-1.51)	-0.353 (-0.27)
FAD	-1.048 (-0.98)	-0.023 (-0.76)	-0.739 (-0.71)	-0.420 (-0.33)	-0.319 (-0.17)
Constant	-11.567*** (-3.77)	-0.297*** (-3.79)	-10.586*** (-3.33)	-10.406*** (-2.84)	-10.328** (-2.23)
Observation	112	112	112	84	56
R-squared coefficient of determination/ goodness of fit	0.296	0.467	0.299	0.316	0.292
F test	0	0	0	0	0
r2_a	0.256	0.436	0.259	0.262	0.205
F value	21.70	20.84	19.34	14.51	10.53

***, **, * means the correlation is significant in 1%, 5%, 10% level respectively; The t value is in parentheses.

3.5 Robustness Test

In order to improve the accuracy of the empirical results, the method of shortening the research interval to 2019-2021 was adopted to test the robustness of the above regression model. The test results are shown in Table 6.

As can be seen from Table 6, the explained variables and explanatory variables of model 1 and model 2 are significantly negatively correlated at the significance level of 1%. The explained variables of Model 3 and Model 4 (lag period 1 and lag period 2, respectively) were significantly positively correlated with explanatory variables at the significance level of 5%. In addition, with the shortening of the research interval, the regression coefficient of explanatory variables increased from 4.472 to 4.901 with the increase in the lag period. Therefore, the empirical model and empirical results constructed in this paper are robust.

Table 6. Robustness Test Results

Explained variable / Explanatory variable	Patent	R&D	Patent (current)	Patent (lag period 1)	Patent (lag period 2)
TAX	-3.658*** (-1.04)	-0.389*** (-2.93)			
R&D			4.266** (1.45)	4.472** (0.90)	4.901** (0.37)
SIZE	0.738*** (4.65)	0.021*** (4.52)	0.662*** (3.77)	0.686*** (3.17)	0.727** (2.26)
YEAR	-0.015 (-0.65)	-0.000 (-0.43)	-0.014 (-0.61)	-0.006 (-0.20)	-0.040 (-0.84)
ROA	-2.441 (-1.40)	-0.1564*** (-2.02)	-1.755 (-0.94)	-1.672 (-0.65)	-0.951 (-0.22)
DAR	-2.476*** (-2.72)	-0.157*** (-5.14)	-1.777 (-1.74)	-2.589* (-1.86)	-1.166 (-0.61)
FAD	-1.257 (-1.05)	-0.025 (-0.72)	-0.967 (-0.83)	-0.703 (-0.48)	-1.409 (-0.51)
Constant	-12.445*** (-3.38)	-0.294*** (-3.01)	-11.818*** (-3.22)	-12.223*** (-2.75)	-12.826* (-1.98)
Observation	84	84	84	56	28
R-squared (coefficient of determination/goodness of fit)	0.317	0.455	0.319	0.338	0.307
F test	0	0	0	0	0
r2_a	0.264	0.413	0.266	0.257	0.209
F Value	24.78	15.96	20.06	13.54	10.03

Note: 1. ***, **, * represent at a significant level of 1%, 5% and 10%, respectively; 2. The t value is in parentheses.

The following conclusions can be drawn by combining Table 5 and Table 6:

(1) In Model 1, the regression coefficient of the corporate income tax burden (Tax) is -4.254 and significant, which is robust. This indicates that the smaller the tax burden of enterprise income tax is, or the higher the level of tax incentives enjoyed by enterprises, will lead to higher innovation performance of enterprises. Thus, hypothesis 1 is valid

(2) In Model 2, the regression coefficient of the corporate income tax burden (Tax) is -0.385, which is significant and robust, indicating that the intensity of R&D investment increases with the increase of the level of tax incentives enjoyed by enterprises. Thus, hypothesis 2 is valid.

(3) In Model 3, the regression coefficient of R&D investment intensity (R&D) is 5.190 and significant, which is robust, indicating that the improvement of R&D

investment intensity will improve the innovation performance of the enterprise in the current period. Thus, hypothesis 3 is true.

(4) In Model 4, R&D investment intensity (R&D) has a significant positive correlation with the innovation performance in the first and second lagging periods and is robust. Thus, hypothesis 4 is valid.

3.6 Analysis of Research Results

a) Tax incentives have a significant positive impact on the innovation performance of pharmaceutical manufacturing enterprises

Tax incentives play an important role in improving the innovation performance of pharmaceutical manufacturing enterprises. Meanwhile, the research results show that innovation performance is also affected by enterprise size, debt-to-assets ratio and other factors. Tax incentives can be differentiated according to the actual business situation of enterprises, to continuously improve the implementation effect of tax incentives on the innovation performance of pharmaceutical manufacturing enterprises.

b) Tax incentives have a significant impact on the R&D investment intensity of pharmaceutical manufacturing enterprises

Tax incentives play a positive role in improving the R&D investment intensity of pharmaceutical manufacturing enterprises. Through additional deduction of R&D expenses, reduction of corporate income tax, investment credit and other tax incentives, enterprises have more after-tax profits and external financing. Sufficient surplus funds guarantee the investment in R&D and innovation activities of enterprises, and reduce the cost and risk of R&D capital.

c) R&D investment intensity has the mediating effect between tax incentives and innovation performance of pharmaceutical manufacturing enterprises

Tax incentives can compensate for the positive externalities of enterprise innovation, reduce research and development costs, guide enterprises to increase investment in innovative research and development, so that enterprises can have more funds to build more professional scientific research teams, improve research and development efficiency, and improve the innovation performance of pharmaceutical manufacturing enterprises. Therefore, we should further improve the tax incentives for R&D and innovation activities.

d) R&D investment intensity has a long-term influence on the innovation performance of pharmaceutical manufacturing enterprises

The influence of R&D investment intensity on the innovation performance of pharmaceutical manufacturing enterprises has a time lag. Innovative R&D activities have strong periodicity and high input in the early stage. Tax incentive policies can be formulated for the early and intermediate stages of R&D innovation to guide pharmaceutical manufacturing enterprises to increase the intensity of R&D investment in order to continuously improve innovation performance.

4. Conclusion

On the basis of science, operability and comparability, this paper puts forward four theoretical hypotheses based on the research results of previous scholars on the theme of the impact of tax incentives on corporate innovation performance. The multiple linear regression model of variables such as tax incentives, R&D investment intensity, enterprise size, establishment time and innovation performance, using STATA software for empirical analysis, draws reliable conclusions about the impact path and effect of tax incentives on innovation performance of Chinese pharmaceutical manufacturing enterprises.. The results show that tax incentives have a significant positive impact on the innovation performance of pharmaceutical manufacturing enterprises. Tax incentives can promote the R&D investment intensity of enterprises, and R&D investment intensity has an intermediary effect between tax incentives and innovation performance, and its effect on innovation performance has a lag. This provides ideas for pharmaceutical manufacturing enterprises to improve their own innovation performance, and provides a reference for the government to introduce tax policies for innovation incentives. This paper selects the data of listed pharmaceutical manufacturing enterprises to study the impact of tax incentives on the innovation performance of enterprises, and the research scope is still relatively limited. In this paper, the number of patent applications is selected as an evaluation index for innovation performance, and in the future, we can start from the aspect of patent quality, establish a more detailed index system, and measure the innovation ability of enterprises more scientifically.

Acknowledgment

This paper is supported by ①The Key Cultivation Base for “The 14th Five-Year Plan” of Educational and Scientific Research (Lifelong Education Research Base(Fundamental Theory Area))(grant no.XJK22ZDJD58)) in Hunan Province.
②Hunan Province colleges and universities’ teaching reform research project (Hunan Provincial Department of education [2019] No.291-Item No.867).③Hunan Provincial Federation of Social Sciences Project No. XSP2023GLC105.④the Chenzhou social science planning project No. Czsskl2022103.

References

[1] Hung-Jen Wang,Peter Schmidt. One-Step and Two-Step Estimation of the Effects of Exogenous Variables on Technical Efficiency Levels.Journal of Productivity Analysis.2002,18(2):241-53.
[2] Michael Peneder. The problem of private under-investment in innovation: A policy mind map. Tech-novation.2008,28(8):518-30.
[3] Hongxiang Tang, Yinchang Li .Tax incentives and enterprise performance: Moderating effects of business environment and enterprise nature..Taxation Research.2020(12):115-21.
[4] Xinghui, Wang Fei, Guo Huili,Zhang Jinhui, Gu Jiangning.Tax incentives, enterprise heterogeneity and substantive innovation: Micro data from listed companies in China.Friends of Accounting.2021(10):22-9.
[5] Fabiani S, Sbragia R. Tax Incentives for Technological Business Innovation in Brazil：The Use of the Good Law-Lei do Bem[J]. Journal of Technology Management & Innovation, 2014, 9(4)：53-63.
[6] Andrew C,Chang. Tax policy endogeneity:evidence from R&D tax credits..Economics of Innovation and New Technology,2018.

[7] Jingyi,Li, Zhenyang Wang, Wu Xianyun.The interaction between policy incentives and R&D investment on innovation performance.Scientific Research Manageme.2020,41(05):99-110.
[8] Congrong Li, Mengfei Jia, Qing Han. The influence of fiscal policy on the innovation performance of military-civilian integration enterprises--Analysis on the mediating effect of innovation input. Scientific and Technological Progress and Measures.2021,38(6):113-21.
[9] Kang He, Yunchen Wang, Liguang Zhang, Limei Wan . Tax Incentives, Innovation Output and Innovation Efficiency——An Empirical Test Based on the Super Deduction Policy of R&D Expenses . East China Economic Management, 2020, 34(1):37-48.
[10] Songbin Wu, Huidan Huang , Kai Zhang. Effectiveness and Influencing Factors of R&D Tax Incentives —Based on 15% Tax Preferential, R&D Investment and Enterprise Innovation Performance—Evidence from Industrial Enterprises and R&D Super Deduction Policy Empirical Comparative Analysis. Science and Technology Progress and Countermeasures,2019,36(11):117-24.
[11] Junjiao Liang, Yuxi Jia. The Impact of Corporate Income Tax Preferences on Enterprise Innovation——An Empirical Analysis Based on Panel Data of Listed Companies . Journal of Central University of Finance and Economics, 2019(09):13-23.
[12] Fenfen Miao . Research on preferential corporate income tax policies to promote innovation investment in the pharmaceutical manufacturing industry. Shandong University of Finance and Economics, 2021.
[13] Yuchen Xiong, Yinguo Li . Research on the relationship between innovation input and innovation output of high-tech enterprises based on canonical correlation analysis . Economic Research Guide, 2019, No.392(06):7-9.
[14] Dianchun Jiang, Xiaowang Pan. The Impact of Digital Economy Development on Enterprise Innovation Performance——Based on the Empirical Evidence of Listed Companies in my country. Journal of Shanxi University (Philosophy and Social Science Edition), 2022,45(01):149-60.
[15] Qiulai Zhang, Yaowei Li. Research on the Impact of Tax Incentives on Enterprise Innovation Performance--An Empirical Analysis of Listed Enterprises in the GEM Manufacturing and Information Technology Industries . Science and Technology Entrepreneurship Monthly, 2018, 31(03): 32-4.
[16] Mengyun Han. The impact of tax incentives on the innovation performance of high-tech enterprises . Southwest University, 2021.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230177

Dynamic Risk Measure for China's Short-Term International Capital Flows

Hui Ma[a, b,1] and Yan Zheng[a]
[a] School of Economics, Harbin University of Commerce, P.R.China
[b] Department of Statistics, University of California Riverside, USA

Abstract. This paper adopts the value-at-risk (VaR) method as the risk measurement framework, and establishes a series of VaR-GARCH models to measure the dynamic risk of short-term international capital random fluctuations. Then, the accuracy of the empirical results is measured by the Backing-Test. The results show that the measurements of VaR-GARCH(1,1)and VaR-GARCH-M models are relatively conservative; the measurement results of VaR-EGARCH model are closest to the true value of short-term international capital flows.

Keywords. Short-term international capital flows, GARCH cluster models, Value at Risk (VaR)

1. Introduction

Excessive short-term international capital inflows and outflows are widely considered to be the main factors leading to inflation and violent fluctuations in asset prices. Especially when China's financial market is underdeveloped and the banking system is fragile, large-scale capital inflows will only trigger excessive investment in the short term and bring about false economic prosperity, but it will plant the seeds of a large number of capital reversals and trigger a currency crisis. Therefore, in today's large-scale and high-speed flow of short-term international capital, how to effectively control its risks and prevent financial crises has become a macroscopic reality that China is facing.

This paper will study the connotation of short-term international capital flow and estimate its flow scale; in terms of practical application, the capital market needs an analysis method that can quantitatively and systematically measure the risks generated by short-term international capital flow. The research on the risk measurement of short-term international capital flow will help to provide more applicable measurement methods for practical work in related fields.

First, the issue of short-term international capital flows is relatively complex, and the division results of different connotations can be obtained from different perspectives, and different connotations involve different economic performances. According to the research focus of this paper, it is mainly to reflect the inflow and outflow of formal channels, relatively fast, and relatively large fluctuations of transnational capital, of course, it also includes the capital that is not directly reflected in the balance of

[1] Correspongding Author, Hui Ma, School of Economics, Harbin University of Commerce, P.R.China, Email: mahui_huc@163.com

payments and is outside the scope of conventional observation. , more speculative cross-border capital flows. Therefore, combining the term method (Cuddington, 1986)[1] with the investor intention judgment method (World Bank, 1985)[2], the short-term international capital flows referred to in this paper include the short-term international capital flows in the balance of payments and off-balance-sheet or hidden capital flows. The caliber of this estimation is relatively wide in the comparison of many methods, which is mainly due to the availability of data (mahui, 2014)[3]. That is: SICF scale = monthly increase in foreign exchange funds - monthly surplus in goods trade - monthly actual use of FDI.

Secondly, the risk measurement carried out by scholars at home and abroad is mainly based on the risk performance and characteristics of international capital flows, with different emphases, different methods and their own advantages and disadvantages.

Some scholars have measured it from the perspective of liquidity based on the theory of interest rate parity, such as the Edwards-Kahn method, savings-investment correlation test method, and offset coefficient test method. Edwards&Kahn (1985)[4] indirectly explained the degree of restriction of the country's international capital flow according to the degree of economic openness of the country. Wang Xiaochun (2001)[5] estimated the degree of capital mobility in China, Malaysia, Thailand and Indonesia. The main idea of the investment-savings related method (referred to as F-H model) proposed by Feldstein&Horioka (1980)[6] is: when a country ' s government implements strict capital control, the domestic capital demand side cannot attract foreign investment, so the country's capital mainly comes from its own country, which makes the country's domestic savings and investment have a high correlation. But Jansen&Schulze (1993)[7] thought that this method ignores the impact of intertemporal budget constraints on the relationship between domestic investment and savings, which leads to fundamental deviations in the measurement. On this basis Jansen(2000)[8] modified the F-H model. Kouri&Porter(1974)[9] proposed the "offsetting coefficient model". According to the "trilemma", if a country's market is highly open and uses a fixed exchange rate, then the country will lose the independence of monetary policy. Under the influence of international arbitrage capital flows, the operations of the central bank will eventually be offset and thus become invalid. According to this idea, as long as the central bank's domestic assets are offset by changes in foreign assets, it is possible to know whether capital flows and the degree of flow.

Other scholars measured the degree of risk from the perspective of the volatility of capital flows. Gabriele (2000)[10] believed that if the original series is not stable, the standard deviation of its annual percentage change can be used as a variability indicator. This indicator can dynamically reflect changes in capital flows, but the percentage calculation of the original indicator uses period t-1 data as the denominator, and in some countries with underdeveloped statistical work, its default data and unstatistical data make the calculation operability and comparability become worse. Robert(2000) [11] used the standard deviation index of the GDP ratio to compare the four types of capital instabilities in low-income, middle-income and high-income countries. Yu Shanping and Zhang Wenxi (2008)[12] constructed a measurement index system for capital volatility and conducted empirical research on China's situation. Alfaro (2004)[13] et al. adopted a relative indicator: the standard deviation of per capita net capital inflows to calculate capital volatility. In this way, the influence of the different scales of capital flows in different countries due to their different economic scales can be eliminated. The World Bank (2001)[14] regarded the median variance of the scale of capital flows as a measure of the volatility of capital flows. However, since this indicator is an absolute value indicator, comparisons between different countries cannot be made. Li Zeguang et al. (2003)[15] adopted a dimensionless statistical indicator— the coefficient of variation of the scale of capital flows. This method had a good explanatory power, but some scholars (mahui, 2010)[16] had questioned the rationality of the index construction.

The above empirical research focused on the mature theoretical basis abroad, corrected some assumptions, and used statistical methods to make the empirical research more suitable for China's situation and made the conclusions more accurate. In addition, short-term international capital flows into the host country and enter different financial institutions to form different financial assets; in terms of the source of risk, short-term international capital flows will form different risks; in terms of specific measurement, none of the risk measurement in previous studies provided appropriate quantitative methods to measure specific risk factors. Therefore, this paper hopes to find a comprehensive measurement method that can measure different financial assets and different risk factors.

In the framework of risk management, quantify qualitative risks to measure the frequency of risk events and the degree of loss that risk events may cause, so that risk managers can achieve appropriate cost risk control under budgetary constraints. The statistical characteristics of market factors, market structure characteristics and volatility rules are one of the foundations of risk measurement and modeling. First, this paper analyzes the related concepts of risk measurement to determine the scope and connotation of SICF risk measurement. Secondly, the value-at-risk (VaR) risk measurement method is used as the measurement framework, and a series of GARCH cluster-VaR models are constructed on the basis of the random fluctuation process to empirically measure the dynamic risk of short-term international capital flows (hereinafter referred to as SICF). Finally, the empirical results of each GARCH cluster model measurement are tested to determine the optimal SICF risk prediction model through the accuracy of the measurement.

2. Models

2.1 VaR-GARCH Cluster Models

The significant advantage of VaR is based on the probability distribution of random variables, and the position of random variables in the distribution is determined according to the characteristics of random variables, and the concept of this risk measurement is to describe future events based entirely on previous information, and this measure is an absolute form of measure that expresses the future risk profile in monetary units of measure. Since this measurement method is based on the collection of previous information, it has nothing to do with that the financial variable is what kind of financial asset, is traded in what kind of financial institution and with what kind of financial instrument. Therefore, VaR is a comprehensive risk measurement method, which can integrate different types of risks caused by fluctuations in different market factors into a single value . Therefore, the VaR risk measurement method provides a basic measurement framework for the SICF risk measurement in this paper. The definition of VaR is: under a certain confidence level, the maximum possible loss value of the entire asset in a certain period of time in the future due to market fluctuations.

But traditional VaR methods calculate risk by assuming specific distributions of financial variables and the statistical characteristics of those distributions. The defect of this kind of method is that the homoscedasticity of implied variable time change and the static measurement cannot reflect the change of time trend and so on. Therefore, this paper uses the GARCH cluster model to solve the description of heteroscedasticity.

Many studies have shown that when the financial market is subject to some external shocks, such as the adjustment of macroeconomic policies or the instability of the country's political situation, the volatility of the forecast results will be relatively large. That is to say, the error in period t depends on the error in period t-1, not on an independent variable in the equation. For the autocorrelation characteristics of the error term reflected in the financial time series, it can be described by the GARCH cluster model. This type of model can well reflect the characteristics of the size of the error term in the financial market changing over time and can also be quantitatively analyzed the nonlinear dependence characterize.

GARCH(1,1) model

The GARCH(1,1) model is the most commonly used and consists of two parts, namely the conditional mean equation and the conditional variance equation, and its general form is:

$$Y_t = X_t \gamma + u_t \tag{1}$$

$$\sigma_t^2 = \omega + \alpha u_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{2}$$

x_t is $1 \times (k+1)$ exogenous variable vector, γ is $(k+1) \times 1$ coefficient vector, and σ_t^2 represents the conditional variance, which is the forecast variance of the t period based on the information of the t-1 period.

GARCH-M model

Engle, Lilien, and Robins (1987)[17] introduced the GARCH-M model, which was used to describe the relationship between risk and return.

$$Y_t = X_t \gamma + \rho \sigma_t^2 + u_t \tag{3}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \ldots + \alpha_p u_{p-1}^2 \tag{4}$$

Among them: the ρ parameter is equivalent to a measure index of risk degree, which mainly measures the impact of conditional variance on the price of financial assets.

TARCH model

Zakoian (1994)[18] and Glosten, Jagannathan, Runkle proposed the TARCH model. The mean value equation of this model has not changed. In order to measure the different fluctuations brought about by the impact of information in different directions, the conditional variance is changed to:

$$\sigma_t^2 = \omega + \alpha u_{t-1}^2 + \gamma u_{t-1}^2 d_{t-1} + \beta \sigma_{t-1}^2 \tag{5}$$

The size of the conditional variance in period t mainly depends on the conditional variance of the previous period and the impact brought by the random error of the previous period. The impact of the asymmetric effect on the fluctuation can be judged by the size of the coefficient γ. The judgment for γ less than 0 is: the asymmetric effect makes the fluctuation larger, and vice versa. And u_{t-1} represents the direction of the impact. When u_{t-1} is a positive impact, that is, less than 0 means good news, and the impact effect on fluctuations is reflected as α times the impact; on the contrary, when

u_{t-1} is a negative impact, the impact effect on fluctuations Reflected as $\alpha+\gamma$ times the impact.

EGARCH model

In order to reflect the relationship between conditional variance and random error in a more flexible and sensitive way, Nelson (1991)[19] proposed EGARCH, which retains the form of the mean equation and rewrites the conditional variance equation as :

$$\ln(\sigma_t^2)=\omega+\alpha\left|\frac{u_{t-1}}{\sigma_{t-1}}\right|+\gamma\frac{u_{t-1}}{\sigma_{t-1}}+\beta\ln(\sigma_{t-1}^2) \tag{6}$$

Because the equation transforms the dependent variable into a logarithmic form, its value range is non-negative, and the leverage effect is also reflected in the form of an index. The existence of asymmetry only requires that γ is not zero.

PARCH model

Also in order to more clearly reflect the impact of large-scale shocks, Ding et al. (1993)[20] proposed to replace the conditional variance with the conditional standard deviation. The conditional standard deviation equation for the PARCH(1,1) model is in the form:

$$\sigma_t^\delta=\omega+\alpha(\left|u_{t-1}^2\right|-\gamma u_{t-1})^\delta+\beta\sigma_{t-1}^\delta \tag{7}$$

$\delta>0$, when $i=1,2,...r$, $|\gamma_i|\le 1$; when $i>r$, $\gamma_i=0$, $r\le p$. In the PARCH(1,1) model, the parameters we need to estimate include $\delta, \alpha, \beta, \gamma$, where δ represents the degree of influence of information shocks on the conditional standard deviation; γ represents the different amplitudes of fluctuations caused by information shocks in different directions, that is, the degree of asymmetric effect .

Finally, the definition of SICF risk value refers to the definition of financial market VaR[21]: under market fluctuations, the maximum possible value of SICF scale includes both the net inflow of short-term international capital and the net outflow of short-term international capital. It further means that under a certain degree of confidence, the maximum possible net inflow and outflow of short-term international capital flows in a specific period of time in the future. Expressed as:

$$\text{VaR(SICF)}=\mu+\sigma_{t-1}\text{VaR}_{1-\alpha}(Z) \tag{8}$$

$$\text{VaR(SICF)}=\mu+\sigma_{t-1}\text{VaR}_{\alpha}(Z) \tag{9}$$

Here, μ is the conditional mean of SICF, σ_{t-1} is the conditional standard deviation, and VaR(Z) is the quantile under the corresponding probability distribution of the SICF sequence. In the short-term international capital market, a country faces the risks of excessive inflow and outflow of SICF respectively. So the left tail (α) indicates that the SICF time series is a net outflow tail, and the right tail ($1-\alpha$) indicates that the SICF time series is a net inflow tail.

2.2 Back Testing method for Accuracy of Risk Measurement models

The accuracy test of the VaR model is the measurement of the degree of coverage of the VaR model to the actual values. Assumed VaR given under 95% confidence level,The accuracy of the VaR model is that whether the probability of the actual value exceeding VaR is less than 5%.

The accuracy of the VaR model has multiple representations, thus it has a variety of test methods, more widely used method is failure frequency test method proposed by Kupiec (1995)[22]. He assumes the independence of VaR estimates with time, the actual losses exceed VaR estimates recorded as a failure, the actual loss is less than the estimated VaR denoted success, that the following " Hit sequence ":

$$\text{Hit}=\begin{cases}1, & \text{SICF}_t < -\text{VaR}_t \\ 0, & \text{SICF}_t \geq -\text{VaR}_t\end{cases} \tag{10}$$

The binomial results of "failure to observe" represent a series of independent Bernoulli test, that is, the "Hit sequence" should obey the Bernoulli distribution with probability p. Therefore, the null hypothesis is defined as follows:

$$H_0\text{: } \text{Hit}_t : \text{Bernoulli}\ (p) \tag{11}$$

Based on probability theory, we can write the likelihood function of a Bernoulli (p) distribution:

$$L(p)=\prod_{t=1}^{T}(1-p)^{1-Hit_t}\,p^{Hit_t}=(1-p)^{T_0}\,p^{T_1} \tag{12}$$

T is the total length of the Hit sequence; T_1 is the sequence in which the value is the sum of a number of "1"; T_0 is the sequence in which the value is the sum of a number of "0".

Kupiec made the most appropriate test of the null hypothesis is the likelihood ratio test:

$$LR=-2\ln\frac{(1-p)^{T_0}\,p^{T_1}}{\left(1-\frac{T_1}{T}\right)^{T_0}\left(\frac{T_1}{T}\right)^{T_1}} : \chi^2(1) \tag{13}$$

Under the assumption of zero, LR statistic subject χ^2 distribution whose degrees of freedom is 1.

In the significance level p, if the LR is greater than the threshold of χ^2 distribution , the null hypothesis H_0 should be rejected; on the contrary, the null hypothesis should be accepted, that the risk measurement model used is sufficiently accurate. If Kupiec LR test value of a VaR model is smaller or the P value is larger, then the more the null hypothesis H_0 can not be rejected, that shows the higher the accuracy of the risk measurement model.

3. Empirical Analysis

3.1 Model Parameter Estimation and Selection

In the course of the calculation method, data were derived from: Exchange of RMB against the U.S. dollar from the State Administration of Foreign Exchange website; Foreign Exchange from the People's Bank of China website; Difference between Imports and Exports of Goods and the FDI from China Economic Information Network. Start and end time data for January 2000 to December 2022.

Using maximum likelihood estimation method proposed for parameters estimation of GARCH cluster model, the results are shown in Table1. In five GARCH models, the model parameters were estimated under GED distributions. Table1 shows the results of a significant parameter estimates, parameters of PARCH and TARCH model are not significant .

Table 1 GARCH Model Parameters Results

	GARCH(1,1)	GARCH-M	EGARCH
c	473858.4***	399709.8**	0.2456***
α	0.5522**	0.3933**	-0.0973***
β	0.3197*	0.4385**	0.9886***
ρ		0.2096**	
γ			0.0261***

Note: "***" indicates significance level of 1%, "**" represents a significant level of 5%, "*" represents a significant level of 10%.

In the GARCH (1,1) model, $\alpha + \beta$ equals 0.8719, parameter values are in the range of constraints. Sum of α and β is close to 1, which indicates that if conditional variance indicated by correlation between variance and time is impacted at a certain time, impact shock is persistent, so if predicting, sustained impact on the future value of persistent shocks should be taken into account.

In the GARCH-M model, ρ satisfies the assumptions, and is positive. This shows that the larger the risk is that the larger conditional standard deviation is relevant of the larger absolute value of SICF. ρ=0.2096 represents the percentage of SICF risk rises with growth.

In EGARCH model, When t +1 period residual is greater than 0. $\alpha+\gamma$ the impact of new information that is what the log σ_t suffered is $\alpha + \gamma$; When t +1 period residuals is less than 0. σ_t is the subject of interest from new. Table 2 displays the results of parameter estimation, the new information rate on the impact of the two directions σ_t numbers were 0.0712 or 0.1234.

3.2 Empirical Results of Models Measure

After the time variance of SICF obtained by GARCH cluster, VaR can be a measure for SICF. Figure 1 to 3 is the risk measurement performed by the VaR-GARCH model. In general, the measurement results of VaR-GARCH are relatively close to the actual value of SICF. Due to the long time series and the large span of index values, it is difficult to visually judge the measurement accuracy of different models from the graph, and it is difficult to choose a better risk measurement model.

Table 2 shows statistical description of VaR of SICF outflow and inflows. From Table 2, you can see the outflows comparison of the three GARCH models. In terms of average, VaR absolute value of GARCH (1,1) is the maximum; standard deviation of VaR value is the maximum in GARCH(1,1) model, standard deviation of VaR value is the minimum in EARCH model. This suggests, what SICF risk measure based on the GARCH cluster model, estimation results of GARCH (1,1) are conservative, estimation results of EGARCH are more optimistic.

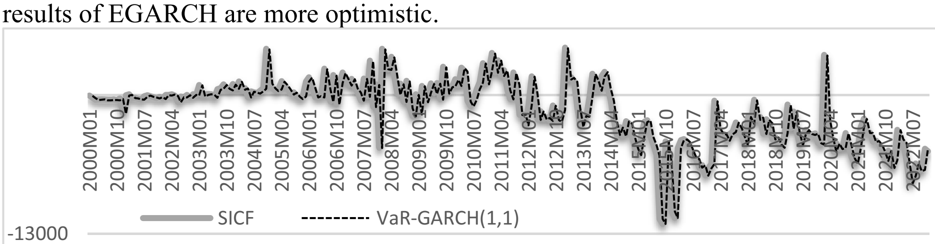


Fig.1 VaR-GARCH(1,1) model for SICF risk measurement

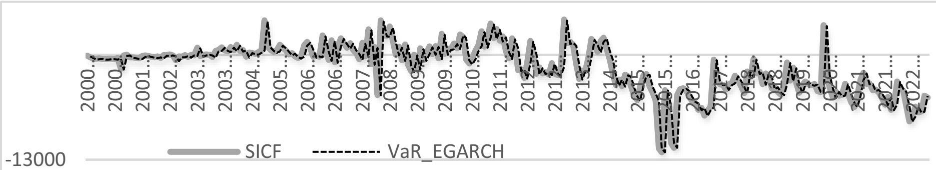


Fig.2 VaR-EGARCH model for SICF risk measurement

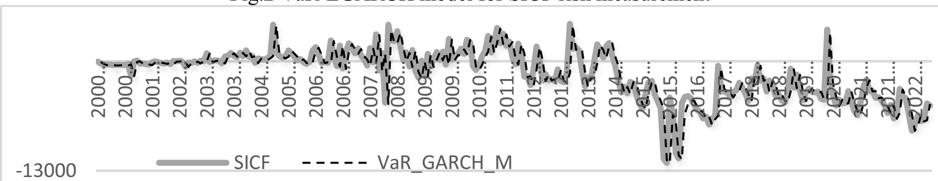


Fig.3 VaR-Garch_M model for SICF risk measurement

In terms of the mean value comparison of VaR of net SICF inflow, the VaR value estimated by GARCH-M is the maximum, the VaR value estimated by EGARCH is the minimum; standard deviation of VaR estimated by GARCH(1,1) is the maximum, standard deviation of VaR estimated by EGARCH is the minimum. This shows that the results based on GARCH-M model estimation are more conservative, results based on EGARCH model estimation are more optimistic.

Table 2 Statistical Description of VaR of the Net Outflow and the Net Inflow (α=95%)

	Net Outflow		**Net Inflow**	
	Mean	Std. Dev.	Mean	Std. Dev.
GARCH(1,1)	-1783.30	1129.28	2413.61	1580.66
GARCH-M	-1489.81	862.26	2524.30	1337.98
EGARCH	-1516.06	685.88	1765.08	885.06

Failure due under the VaR measure based on EGARCH model will be carried out under extreme conditions in the SICF. So when major changes occur in economic situation, using EGARCH model to measure the risk of SICF fluctuations should be cautious; in terms of GARCH (1,1)-VaR and GARCH-M-VaR measure, because in the event of an extreme case of the former measure has a wider range, its measure also tend to be conservative. In the gentle SICF range, EGARCH-VaR results are the closest to the actual SICF values. In positive territory, GARCH-M-VaR measure is more conservatism than GARCH (1,1)-VaR measure ; In negative territory, the situation is opposite. But extreme values appear in SICF, EGARCH-VaR measure failure happens; at this point, whether negative or positive range, GARCH(1,1)-VaR measure should be more conservative than GARCH-M-VaR measure.

3.3 Accuracy Test of Empirical Results

The results of the empirical measure under certain model analysis are carried out under visual observation, do not have specific criteria to measure the accuracy of a specific measure. Therefore, we will use the results of the above test method with "Back Testing". Table3 shows the accuracy test results of VaR measure under the left and right distribution of SICF. The normal font indicates its acceptance of the null hypothesis test, that is considered to be accurate test results; underlined font t indicates their test results reject the null hypothesis, that is considered to be inaccurate test results.

First, comparing from different significance level. For high significance level such as 1%, some test results accept the null hypothesis, under 5% and 10% significance level, all tests have rejected the results. It means that when the tail distribution is wide, when the greater the probability of extreme events occur, as excessive SICF outflow occurred many times, various tests appeared measure fails.

Second, LR values of EGARCH-VaR measure in 10% level of significance are significantly greater LR values in 5% significance level(in refusing domain), this shows that the accuracy of VaR measurement is declining as the significant level.

Table 3 Back Testing (Left and Right)

α		GARCH(1,1)		EGARCH		GARCH-M	
		Left	**Right**	**Left**	**Right**	**Left**	**Right**
1%	Failure number	0	0	2	2	0	0
	LR	-∞	-∞	0.22	0.22	-∞	-∞
5%	Failure number	0	0	2	5	0	0
	LR	-∞	-∞	5.2485	0.6953	-∞	-∞
10%	Failure number	0	0	2	11	0	0
	LR	-∞	-∞	17.5066	0.813	-∞	-∞

4. Conclusion

This paper defines SICF as the sum of short-term international capital flows in the balance of payments and hidden international capital flows outside the balance sheet, and estimates the size of SICF from January 2001 to December 2022. This paper uses VaR as the basic measurement framework for the risk measurement of SICF to measure the maximum loss value that may occur in a certain period in the future, and combines the GARCH cluster model to reflect the nonlinear change of SICF risk.

The results of the model parameter estimation show that VaR-GARCH(1,1), VaR-EGARCH and VaR-GARCH-M models are significant, VaR-TARCH and VaR-PARCH models are not significant. In the VaR-GARCH (1, 1) model, $\alpha+\beta<1$ indicates that the shock has a continuous impact on the future; in the VaR-EGARCH model, $\rho<0$ indicates that greater risk increases with larger value of SICF; in the EGARCH model, the impact of information is asymmetric, and the risk of outflow from SICF is greater than the risk of inflow. In practice, risk managers should continue to monitor the impact of SICF in the medium and long term, and price the medium and long-term impact in the risk pricing of assets; set the SICF early warning range, and focus on monitoring the ultra-volatile range, in which the short-term international capital outflow risks are greater than inflows risks so that more attention should be paid to outflow risk monitoring and risk valuation.

According to the results of Backing-Test under different significance, measure of VaR-GARCH(1,1) and VaR-GARCH-M model is more conservative. It is found in the empirical results, in some extreme cases when the measure used by VaR-EGARCH model fails, both models measure is still valid. But in general case, we also found that the two measurement results of the model are significantly conservative than measure results of VaR-EGARCH model. Measure results of VaR-EGARCH model are closest to the true value of SICF. However, in the extreme case of the measure, the failure cases of models are increasing.That is to say, if the accuracy of risk measurement is the primary consideration of SICF risk managers, then the VaR-EGARCH model is a better choice; if the reliability of risk measurement is the first choice of SICF risk managers, then VaR-GARCH(1, 1) and VaR-GARCH-M model are better choices.

However, it needs to be emphasized that in extreme cases, the number of measurement failures of the three SICF risk measurement models increases significantly.

References

[1] Cuddington. Capital Flight: Estimate, Issues, and Explanations. Princeton Studies in International Finance, 1986, 58 (5):1-40.

[2] The World Bank. World Debt Tables, Washington, 1985:1-56.

[3] Ma Hui. Short-term international capital flows risk measure based on coherent risk measure, International Conference on Construction and Real Estate Management, 2014, 7:832-838.

[4] Edwards S,Khan M S. Interest rate determination in developing countries:A conceptual framework. IMF Staff Papers, 1985,9

[5] Wang Xiaochun. Capital mobility estimation method and its application in developing countries. World Economy, 2001(7):20-26.

[6] Feldstein M, Horioka C. Domestic saving and international capital flows. Economic Journal, 1980, 90(12):314-329.

[7] Jansen, W. Jos; Schulze, Gunther G. The Effects of Capital Controls on Capital Mobility - An Empirical Investigation for Norway, University of Konstanz, Department of Economic Discussion Paper, 1993, series II, forthcoming.

[8] Jansen, W. Jos, International capital mobility: evidence from panel data, Journal of International Money and Finance , 2000, 19 , (4), 507-511

[9] P.J.K. Kouri and M.G.Porter. International Capital Flows and Protfolio Equilibrium. Journal of Political Economy, 1974,5(1):341-356.

[10] Gabriele A K Boratav, A Parikh. Instability and volatility of capital flows to developing countries. World Economy, 2000, 10(4):123-134.

[11] Robert Osei, Oliver Morrissey, Robert Lensink. The Volatility of Capital Inflows: Measures and Trends for Developing Countries, CREDIT Research Paper. Centre for Research in Economic Development and International Trade, University of Nottingham, 2000,02/20.

[12] Yu Shanping, Zhang Wenxi. Measurement of the Volatility of China's Non-FDI Capital Inflows. Journal of Southeast University (Philosophy and Social Sciences), 2008, 10(5):14-18.

[13] Alfaro L, A Chanda, S Kalemli-Ozan, S Sayek. Volatility of Capital Flows:Bad Policies or Bad Institutions. Preliminary Version，2004:132-143.

[14] World Bank. Global Development Financing—Analysis and Summary Table. Beijing: China Financial and Economic Press, 2001.

[15] Li Zeguang, Fu Fei, Tang Weixia. Analysis of Capital Volatility and Its Economic Effect. Finance and Economics Research, 2003, 29(12):11-17.

[16] Ma Hui. Economic analysis and empirical study on factors affecting short-term capital flows of China, International Conference on Management Science & Engineering (17th), 2010, 7:1211-1217.

[17] Engle, R. F., Lilien, D. M. and Robins, R. P. Estimating time-VARying risk premia in the term structure: The ARCH-M model. Econometrica, 1987, 55(5):391-407.

[18] Jean-Michel Zakoian. Threshold heteroskedastic models. Journal of Economic Dynamics and Control, 1994, 18(5):931–955.

[19] Nelson, Daniel B. Conditional Heteroskedasticity in Asset Returns: A New Approach. Econometrica, Econometric Society, 1991, 59(2):347-70.

[20] Zhuanxin Ding, Clive W.J. Granger, Robert F. Engle. A long memory property of stock market returns and a new model. Journal of Empirical Finance, 1993, 1(1):83-106.

[21] Jorion, P. Financial Risk Manager Handbook. John Wiley& Sons, 2003:21-41.

[22] Kupiec P. Techniques for verifying the accuracy of risk measurement models. Journal of Derivatives, 1995, 4(2):173-184.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230178

New Type of Chaotic Attractors and Their Applications

Inese BULA [a,b,1], Michael A. RADIN [c]
[a] *Department of Mathematics, University of Latvia*
[b] *Institute of Mathematics and Computer Science of University of Latvia*
[c] *Rochester Institute of Technology, School of Mathematical Sciences, Rochester, New York 14623, U.S.A.*

Abstract. We study the periodic solutions of a discrete neuron model for period two or period three of the parameter (internal decay rate) $(\beta_n)_{n=0}^{\infty}$. The novelty of this research is finding a chaotic attractor for certain interval, outside the defined interval the solution goes to positive infinity or to negative infinity. The investigation can be useful in the design of chaos-based neural networks architecture.

Keywords. neuron model, discrete dynamical system, difference equation, chaotic attractor

1. Introduction

In [1], the authors investigated the delayed differential equation

$$x'(t) = -g(x(t-\tau)), \tag{1}$$

that is used to model a single neuron, where $g\colon \mathbf{R} \to \mathbf{R}$ is signal function and $\tau \leq 0$ is a synaptic transmission delay. A more historical insight into this equation can be found in [2]. From Eq. (1) a model for a single neuron is obtained $x'(t) = -g(x[t]))$, where $[t]$ denote a greatest integer function. When we integrate this equation from n to $t \in [n, n+1]$ we get $x(t) = x(n) - \int_n^t g(x([s]))\,ds = x(n) - g(x(n))(t-n)$. By letting $t \to n+1$ and denoting $x(n) = x_n$ a difference equation is obtained $x_{n+1} = x_n - g(x_n)$. This equation is generalized for a discrete-time network of a single neuron model ([3]):

$$x_{n+1} = \beta x_n - g(x_n),\ n = 0, 1, 2, ..., \tag{2}$$

where $\beta > 0$ is an internal decay rate and the signal function g is the following piecewise constant function with McCulloch-Pitts nonlinearity:

$$g(x) = \begin{cases} 1, x \geq 0, \\ -1, x < 0. \end{cases} \tag{3}$$

[1]Corresponding Author: Inese Bula, Department of Mathematics, Faculty of Physics, Mathematics and Optometry, University of Latvia, Jelgavas 3, Riga, and Institute of Mathematics and Computer Science of University of Latvia, Raina boulv. 29, Riga, Latvia; E-mail: inese.bula@lu.lv

Several authors investigated Eq. (2) ([3,4,5,6,7,8], etc.). Difference equations have been used as mathematical models for applications including neurons (see [9]).

The novelty of the authors of this article is the proposal to view the internal decay rate β as a sequence of periodic numbers $(\beta_n)_{n=0}^{\infty}$. In [10,11,12], the authors investigated the periodic solutions of a discrete neuron model when $(\beta_n)_{n=0}^{\infty}$ is periodic with periods two and three. The existence of periodic points is different for sequences with period two (even number) and three (odd number). In [13], the authors consider the situation when the sequence is periodic with period 2 and show that at certain values of the coefficients β_0 and β_1 a chaotic attractor is formed. It could be said that we deliberately create such data (big data) that could be used to achieve certain goals.

In this article, we will focus on the sequences with period 3. We study the following non-autonomous piecewise linear difference equation:

$$x_{n+1} = \beta_n x_n - g(x_n), \tag{4}$$

where $(\beta_n)_{n=0}^{\infty}$ is a periodic sequence with period three where

$$\beta_n = \begin{cases} \beta_0, & \text{if } n = 3k, \\ \beta_1, & \text{if } n = 3k+1, \; k = 0,1,2,..., \\ \beta_2, & \text{if } n = 3k+2, \end{cases}$$

$\beta_0 > 0$, $\beta_1 > 0$, $\beta_2 > 0$, all β_i, $i = 1,2,3$, are not equal, with function g in form Eq. (3).

If we consider the right side of difference Eq. (4) as a function $h : \mathbf{R} \to \mathbf{R}$ and let $x_n = h^n(x_0)$, $x_0 \in \mathbf{R}$, $n = 1,2,...$, then we obtain the first order difference equation $x_{n+1} = h(x_n)$ with initial condition $x_0 \in \mathbf{R}$. From the definition of Eq. (4), it follows that first iteration of function h is in the form: $x_1 = h(x_0) = \begin{cases} \beta_0 x_0 - 1, & x_0 \geq 0, \\ \beta_0 x_0 + 1, & x_0 < 0. \end{cases}$ Depending on the circumstance, sometimes it is more convenient to describe the dynamics more easily with the behavior of a function, and at other times with a difference equation.

In general, we consider a first order difference equation (see [14]) $x_{n+1} = f(x_n)$, where $f : \mathbf{R} \to \mathbf{R}$. Then the *orbit of a point* $x_0 \in \mathbf{R}$ is defined to be the set of points $\{x_0, x_1 = f(x_0), x_2 = f(f(x_0)) = f^2(x_0), ..., x_n = f^n(x_0), ...\}$. A point x^* is said to be a *fixed point* of the map f or an *equilibrium point* of equation $x_{n+1} = f(x_n)$ if $f(x^*) = x^*$.

The concept of periodicity is one of the most important notion in the field of dynamical systems. Its importance follows from the fact that many physical phenomena have certain patterns that repeat themselves (for example, the motion of a pendulum, the motion of planets, the population size of blowflies or other insects at time n, the price of commodity at time n).

Let $\bar{x}$ be in the domain of a mapping f. A point $\bar{x}$ is said to be a *periodic point* of f with period k if $f^k(\bar{x}) = \bar{x}$ for some positive integer k. Note that $\bar{x}$ is a periodic point with period k if it is a fixed point of the map f^k.

We organize our paper as follows. In the next section we present results about Eq. (4). In section 3 we analyze the Lyapunov exponent and find this exponent for our dynamical system. We show that for certain values of coefficients β_i there exists a chaotic attractor. At the end we give some concluding remarks, applications and future ideas.

2. Some results about difference equations with period three coefficients

In this chapter, we refer to some of the most important results of article [11].

We remark that Eq. (4) with g in form Eq. (3) has no equilibrium points.

Theorem 1 *The equation (4) has no periodic orbits with periods* $3n+1$ *and* $3n+2$, $n = 0, 1, 2, \ldots$.

The number of periodic orbits depends on the relationship between the parameters β_0, β_1 and β_2. In case with two periodic coefficients what both less than 1 exist only periodic solution with period two but in case with three periodic coefficients what all three are less than 1 not exist periodic solution with period three, in this case exist only periodic solution with period six.

If the product of the coefficients $\beta_1\beta_2\beta_3$ is strictly greater than 1 then there are always solutions with period three.

Theorem 2 *If* $\beta_0\beta_1\beta_2 > 1$ *then initial conditions*

$$x_0 = \frac{\beta_1\beta_2 + \beta_2 + 1}{\beta_0\beta_1\beta_2 - 1} \text{ and } x_0 = -\frac{\beta_1\beta_2 + \beta_2 + 1}{\beta_0\beta_1\beta_2 - 1}$$

form periodic solutions of equation (4) with period three; all points of orbit are positive in first case and negative in second case and both orbits are unstable.

If $\beta_0\beta_1\beta_2 > 1$, then we have observed that our difference equation exhibits unbounded solutions. Observe that in the conditions of the next theorem, the inequalities include the initial points of a cycle with period three (Theorem 2).

Theorem 3 *If* $\beta_0\beta_1\beta_2 > 1$ *and* $x_0 > \dfrac{1 + \beta_2 + \beta_1\beta_2}{\beta_0\beta_1\beta_2 - 1}$, *then* x_0 *forms unbounded solutions of Eq. (4) - going to* $+\infty$. *If* $\beta_0\beta_1\beta_2 > 1$ *and* $x_0 < -\dfrac{1 + \beta_2 + \beta_1\beta_2}{\beta_0\beta_1\beta_2 - 1}$, *then* x_0 *forms unbounded solutions of Eq. (4) - going to* $-\infty$.

3. Chaotic attractor

In this chapter, we will prove that the Eq. (4) (we can also say the function h) forms a chaotic attractor in the set $[-1, 1]$ under the condition $1 < \beta_i \le 2$, $i = 0, 1, 2$.

First, our aim is to determine the invariant interval (a set $I \subset X$ is said to be invariant under the map $f : X \to X$ if $f(I) = I$).

The invariant interval must contain the entire interval $[-1, 1]$.

Now suppose that $x_0 \in [-1, 1]$. Then the following statements hold true:

a) if $0 \le x_0 \le 1$, then $-1 = 0 - 1 \le h(x_0) = \beta_i x_0 - 1 \le 2 \cdot 1 - 1 = 1$, $i = 0, 1, 2$,

b) if $-1 \le x_0 < 0$, then $-1 = 2 \cdot (-1) + 1 \le h(x_0) = \beta_i x_0 + 1 < 0 + 1 = 1$, $i = 0, 1, 2$.

Definition 1 *The Lyapunov exponent* $\lambda(x_0)$ *of the orbit* $\{x_0, x_1, x_2, \ldots\}$ *is defined as*

$$\lambda(x_0) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \ln |f'(x_k)|,$$

provided that the limit exists.

In [14], the authors showed that if the Lyapunov exponent $\lambda > 0$, then the sensitivity dependence on initial conditions exists. The Lyapunov exponent at a point x measures the growth in error per iteration. As the Lyapunov exponent becomes larger, the magnification of error becomes greater.

Theorem 4 *If $\beta_0\beta_1\beta_2 > 1$, then function h have a positive Lyapunov exponent.*

Proof. For every $\beta_0 > 0$, $\beta_1 > 0$, $\beta_2 > 0$ and arbitrary initial point x_0 (what is not point of discontinuity) the Lyapunov exponent is

$$\lambda(x_0) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \ln|h'(x_k)| =$$

$$= \lim_{n\to\infty} \frac{1}{n} (\ln\beta_0 + \ln\beta_1 + \ln\beta_2 + \ldots + \ln\beta_0 + \ln\beta_1 + \ln\beta_2 + i_1 \cdot \ln\beta_0 + i_2 \cdot \ln\beta_1) =$$

$$= \lim_{n\to\infty} \frac{1}{n} \left(\frac{\ln(\beta_0\beta_1\beta_2)\cdot(n-i_1-i_2)}{3} + i_1 \cdot \ln\beta_0 + i_2 \cdot \ln\beta_1 \right) = \frac{\ln(\beta_0\beta_1\beta_2)}{3} > \frac{\ln 1}{3} = 0,$$

where

$$i_1 = \begin{cases} 0, \text{ if } n = 3m, \\ 1, \text{ if } n = 3m+1 \text{ or } 3m+2, \end{cases} \quad i_2 = \begin{cases} 0, \text{ if } n = 3m \text{ or } 3m+1, \\ 1, \text{ if } n = 3m+2, \end{cases} \quad m \in \mathbf{N}.$$

■

We will show that for certain values of the coefficients β_1, β_2 and β_2 Eq. (4) forms a chaotic system.

A Discrete Dynamical System, denoted by DDS for short, is the description of an evolutive phenomenon in terms of a map f whose image is contained in its domain X. Then the pair $\{X, f\}$ is called DDS.

Definition 2 *([15], see [16]) A set $A \subset I$ is called an attractor for a DDS $\{I, f\}$ if the following conditions hold:*
1) A is closed;
2) A is invariant;
3) there exists $\eta > 0$ such that, for any $x \in I$ fulfilling $dist(x,A) < \eta$, we have $\lim_{k\to\infty} dist(f^k(x),A) = 0$;
4) A is a minimal, that is there are no proper subsets of A fulfilling 1), 2) and 3).

Definition 3 *([15]) If A is an attractor of function f, then the set*

$$\left\{ x \in \mathbf{R} | \lim_{k\to\infty} f^k(x) \in A \right\}$$

is called an attraction basin of attractor A.

Definition 4 *([16]) An invariant set A is called a chaotic attractor provided it is an attractor and f has sensitive dependence on initial conditions on A (or f have a positive Lyapunov exponent on A).*

Theorem 5 *Let* $1 < \beta_0 \le 2,\ 1 < \beta_1 \le 2,\ 1 < \beta_2 \le 2$ *and at least two from them are different. Then* $[-1, 1]$ *is a chaotic attractor of function* h *and attraction basin is* $\left]-\frac{\beta_1\beta_2+\beta_2+1}{\beta_0\beta_1\beta_2-1}, \frac{\beta_1\beta_2+\beta_2+1}{\beta_0\beta_1\beta_2-1}\right[$.

Proof. In case $1 < \beta_0 \le 2, 1 < \beta_1 \le 2, 1 < \beta_2 \le 2$ and at least two from them are different, the interval $[-1, 1]$ is an invariant set for the function h and the Lyapunov exponent is positive by Theorem 4. Thus $[-1, 1]$ is a chaotic attractor of function h.

Since $1 < \beta_0 \le 2, 1 < \beta_1 \le 2, 1 < \beta_2 \le 2$ and

$$\beta_1\beta_2 + \beta_2 + 1 \ge \beta_0\beta_1\beta_2 - 1 \Leftrightarrow 2 \ge \beta_0\beta_1\beta_2 - \beta_1\beta_2 - \beta_2 = \beta_2(\beta_1(\beta_0 - 1) - 1)$$

then $\frac{\beta_1\beta_2+\beta_2+1}{\beta_0\beta_1\beta_2-1} \ge 1$. Similarly, it can be proved that all points of a cycle with period three are greater than 1, that is, also $\frac{\beta_0\beta_2+\beta_0+1}{\beta_0\beta_1\beta_2-1} \ge 1$ and $\frac{\beta_0\beta_1+\beta_1+1}{\beta_0\beta_1\beta_2-1} \ge 1$.

Let $\frac{\beta_1\beta_2+\beta_2+1}{\beta_0\beta_1\beta_2-1} > 1$. Our aim is to show that for all

$$x_0 \in \left]-\frac{\beta_1\beta_2 + \beta_2 + 1}{\beta_0\beta_1\beta_2 - 1}, \frac{\beta_1\beta_2 + \beta_2 + 1}{\beta_0\beta_1\beta_2 - 1}\right[\setminus [-1, 1]$$

the orbit by the function h eventually falls in the interval $[-1, 1]$. We consider only the case when $1 < x_0 < \frac{\beta_1\beta_2+\beta_2+1}{\beta_0\beta_1\beta_2-1}$. The case when $-\frac{\beta_1\beta_2+\beta_2+1}{\beta_0\beta_1\beta_2-1} < x_0 < -1$ is similar and will be omitted.

If $1 < x_0 < \frac{\beta_1\beta_2+\beta_2+1}{\beta_0\beta_1\beta_2-1}$, then

$$0 < \beta_0 - 1 < x_1 = \beta_0 x_0 - 1 < \frac{\beta_0(\beta_1\beta_2 + \beta_2 + 1)}{\beta_0\beta_1\beta_2 - 1} - 1 = \frac{\beta_0\beta_2 + \beta_0 + 1}{\beta_0\beta_1\beta_2 - 1}.$$

If $0 < x_1 \le 1$, then the proof is complete. If this is not the case, then $1 < x_1 < \frac{\beta_0\beta_2+\beta_0+1}{\beta_0\beta_1\beta_2-1}$ and therefore

$$0 < \beta_1 - 1 < x_2 = \beta_1 x_1 - 1 < \frac{\beta_1(\beta_0\beta_2 + \beta_0 + 1)}{\beta_0\beta_1\beta_2 - 1} - 1 = \frac{\beta_0\beta_1 + \beta_1 + 1}{\beta_0\beta_1\beta_2 - 1}.$$

If $0 < x_2 \le 1$, then the proof is complete. If this is not the case, then $1 < x_2 < \frac{\beta_0\beta_1+\beta_1+1}{\beta_0\beta_1\beta_2-1}$ and therefore

$$0 < \beta_2 - 1 < x_3 = \beta_2 x_2 - 1 \frac{\beta_2(\beta_0\beta_1 + \beta_1 + 1)}{\beta_0\beta_1\beta_2 - 1} - 1 = \frac{\beta_1\beta_2 + \beta_2 + 1}{\beta_0\beta_1\beta_2 - 1}.$$

Provided that $x_n \notin [-1, 1]$, by induction we then conclude that

$$\begin{aligned}
&1 < x_{3k} = \beta_0^k\beta_1^k\beta_2^k x_0 - \beta_0^{k-1}\beta_1^k\beta_2^k - \beta_0^{k-1}\beta_1^{k-1}\beta_2^k - \ldots - \beta_2 - 1 < \tfrac{\beta_1\beta_2+\beta_2+1}{\beta_0\beta_1\beta_2-1},\\
&1 < x_{3k+1} = \beta_0^{k+1}\beta_1^k\beta_2^k x_0 - \beta_0^k\beta_1^k\beta_2^k - \beta_0^k\beta_1^{k-1}\beta_2^k - \ldots - \beta_0 - 1 < \tfrac{\beta_0\beta_2+\beta_0+1}{\beta_0\beta_1\beta_2-1},\\
&1 < x_{3k+2} = \beta_0^{k+1}\beta_1^{k+1}\beta_2^k x_0 - \beta_0^k\beta_1^{k+1}\beta_2^k - \beta_0^k\beta_1^k\beta_2^k - \ldots - \beta_1 - 1 < \tfrac{\beta_0\beta_1+\beta_1+1}{\beta_0\beta_1\beta_2-1},\\
&k = 0, 1, 2, \ldots.
\end{aligned}$$

Next note that the difference between the iterations x_{3k} and x_{3k+3} is

$$x_{3k}-x_{3k+3}=\beta_0^k\beta_1^k\beta_2^k x_0-(\beta_0^{k+1}\beta_1^{k+1}\beta_2^{k+1}x_0-\beta_0^k\beta_1^{k+1}\beta_2^{k+1}-\beta_0^k\beta_1^k\beta_2^{k+1}-\beta_0^k\beta_1^k\beta_2^k)=$$
$$=(\beta_0\beta_1\beta_2)^k(\beta_0\beta_1\beta_2-1)\left(\frac{\beta_1\beta_2+\beta_2+1}{\beta_0\beta_1\beta_2-1}-x_0\right), k=0,1,2,\ldots.$$

Since all multipliers are positive and $\lim\limits_{k\to\infty}(\beta_0\beta_1\beta_2)^k=+\infty$ then the difference between x_{3k} and x_{3k+3} increases and we then get $x_0>x_3>x_6>...>x_{3k}>x_{3k+3}>....$ Thus we conclude that there exists $k\in\mathbf{N}$ such that $x_{3k}\le 1$.

Similarly, the difference between iterations x_{3k+1} and x_{3k+4}, x_{3k+2} and x_{3k+5} increases and we get $x_1>x_4>x_7>...>x_{3k+1}>x_{3k+4}>...$ and $x_2>x_5>x_8>...>x_{3k+2}>x_{3k+5}>....$ Thus we conclude that there exists $k_1\in\mathbf{N}$ such that $x_{3k_1+1}\le 1$ and there exists $k_2\in\mathbf{N}$ such that $x_{3k_2+2}\le 1$. ■

Example 1 Suppose that $\beta_0=1.92$, $\beta_1=1.26$ and $\beta_2=1.5$. We obtain the following period-3 cycle $\{1.669963481, 2.206329884, 1.779975654\}$ and the following basin of attraction $]-1.669963481, 1.669963481[$. If we start with initial condition $x_0=1.66$ (a point close to the boundary of the interval), then we observe the situation described in Theorem 5, where the first ten iterations of the solution are greater than 1 (see Fig. 1). Then $x_{11}=0.628189142<1$ and all other points of the solution are in the interval $[-1, 1]$. In Fig. 1 we see that

$$x_0>x_3>x_6>x_9>x_{12},\quad x_1>x_4>x_7>x_{10}>x_{13},\quad x_2>x_5>x_8>x_{11}.$$

The behavior of the other points cannot be clearly described, but all other points are located in the invariant interval (attractor) $[-1, 1]$.

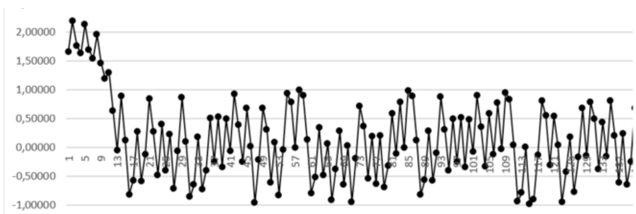


Figure 1. First 140 values of solution of difference equation (4) if $\beta_0=1.92$, $\beta_1=1.26$ and $\beta_2=1.5$, and $x_0=1.66$.

4. Conclusion

In this article, we have shown a new example of a chaotic attractor. The novelty is based on the use of periodicity. It could be that at certain values of the parameters β_i, $i=0,1,2$, which are not all in the interval $]1;2]$, there exist some other chaotic attractors. The case with $\beta_0=1$, $\beta_1=2$ and $\beta_2=3$ mentioned in the article [11] is interesting. Although there could be an invariant interval $[-2, 2]$ here, numerical experiments show that all solutions are periodic or eventually periodic.

The properties of chaos (sensitivity to initial conditions) can be used in random number generation (see, for example, [17]) as well as in cryptography (see, for example, [18]). An important role here is played the uniform distribution of elements of the solution. For example, the histogram of the solution of the difference equation (4) with $\beta_0 = 2$, $\beta_1 = 1.8$ and $\beta_2 = 2$ with the initial condition $x_0 = 0.6$ is shown in Fig. 2. The larger the values of β_i, $i = 0, 1, 2$, the greater the sensitivity to the initial conditions (larger Lyapunov exponent), the more appropriate the histogram looks.

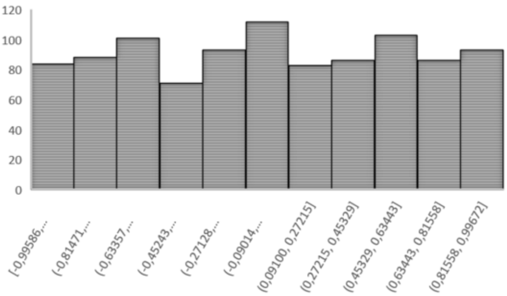


Figure 2. The histogram of the solution (first 1000 values) of the difference equation (4) with $\beta_0 = 2$, $\beta_1 = 1.8$, $\beta_2 = 2$ and initial condition $x_0 = 0.6$.

References

[1] Wu J. Introduction to Neural Dynamics and Signal Transmission Delay. Berlin: De Gruyter; 2001.

[2] Halanay A, Yorke JA. Some New Results and Problems in the Theory of Differential-Delay Equations. SIAM Review. 1971; 13:55-80.

[3] Chen Y. All solutions of a class of difference equations are truncated periodic. Appl. Math. Lett. 2002; 15:975-9.

[4] Bin H, Huang L, Zhang G. Convergence and periodicity of solutions for a class of difference systems. Adv. Difference Equ. 2006; 2006: [1]-10.

[5] Wei Z, Huang L, Meng Y. Unboundedness and periodicity of solutions for a discrete-time network model of three neurons. Appl. Math. Model. 2008; 32:1463-74.

[6] Zhou Z. Periodic orbits on discrete dynamical systems. Comput. Math. Appl. 2003; 45:1155-61.

[7] Zhu H, Huang L. Dynamics of a class of nonlinear discrete-time neural networks. Comput. Math. Appl. 2004; 48:85-94.

[8] Yi T, Zhou Z. Periodic solutions of difference equations. J. Math. Anal. Appl. 2002; 286:220-29.

[9] Ibaz B, Casado JM, Sanjuan MA. Map-based models in neuronal dynamics. Phys. Rep. 2011; 501:1-74.

[10] Bula I, Radin MA. Periodic orbits of a neuron model with periodic internal decay rate. Appl. Math. Comput. 2015; 266:293-303.

[11] Bula I, Radin MA, Wilkins N. Neuron model with a period three internal decay rate. Electron. J. Qual. Theory Differ. Equ. 2017; 46:[1]-19.

[12] Bula I, Radin MA. Eventually periodic solutions of single neuron model. Nonlinear Anal. Model. Control. 2020; 25:903-18.

[13] Bula I, Radin MA. Chaotic single neuron model with periodic coefficients with period two. Unpublished.

[14] Elaydi SN. Discrete Chaos: with Applications in Science and Engineering. Sec.ed., Boca Raton: ChapmanHall/CRC; 2008.

[15] Salinelli E, Tomarelli F. Discrete Dynamical Models. Unitext, 76, Switzerland: Springer; 2014.

[16] Robinson C. Dynamical Systems: Stability, Symbolic Dynamics, and Chaos. Florida: CRS Press; 1995.

[17] Addabbo T, Fort A, Rocchi S, Vignoli V. Digitized chaos for pseudo-random number generation in cryptography. In: Kocarev L, Lian S, editors. Chaos-based Cryptography. Theory, Algorithms and Applications. Berlin, Heidelberg: Springer-Verlag; 2011. p. 67-97.

[18] Schmitz R. Use of chaotic dynamical systems in cryptography. Journal of The Franklin Institute. 2001; 338:429-41.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230179

A Study of the Impact of Excess Cash Holdings on Corporate Innovation

Pan Ma [1]
School of Economics and Management, Lanzhou University of Technology, Lanzhou, Gansu, 730000, China

Abstract: This paper investigates the relationship between excess cash holdings, product market competition and corporate innovation by using an empirical research method with a sample of A-share manufacturing companies listed in Shanghai and Shenzhen in China from 2011 to 2021. The study finds that: manufacturing companies' excess cash holdings have a positive impact on corporate innovation; product market competition positively regulates the relationship between excess cash holdings and corporate innovation, and the more intense the product market competition is, the more significant the regulating effect is. The research in this paper provides a reference for manufacturing firms to optimize corporate cash holdings and promote corporate innovation in different competitive market environments.

Keywords: level of corporate cash holdings; product market competition; corporate innovation

1. Introduction

Enterprise innovation is a topic that has been commonly mentioned in recent years, and innovation in manufacturing enterprises in particular is of paramount importance. There are many factors that affect innovation in manufacturing enterprises, among which the lack of sufficient financial resources is one of the biggest problems that prevent most enterprises from carrying out technological innovation. However, in emerging market countries such as China, where capital markets and financial systems are relatively underdeveloped, there are few opportunities to obtain external financing for R&D. As a result, internal cash holdings are the main source of funding for innovative R&D. Arrow [1]study shows that moral hazard and adverse selection problems due to information asymmetry prevent external financing of technological innovation inputs, and firms then use internal cash holdings. Kamien and Schwartz[2] were the first to test theoretically that firms with a need for innovation increase their cash holdings. In China, Qingquan Tang and Xin Xu[3] showed that R&D investment is dependent on firms' internal capital due to its characteristics, and explored its impact on the relationship between internal capital and investment from the perspective of equity concentration and equity structure. Xin Lu et al.[4] classified high-tech firms listed in 2007-2009 into two categories: cash-rich and cash-deficient, and found that cash-deficient firms have significant cash flow sensitivity of R&D investment, and this relationship is stronger for non-state owned firms and small-sized firms. But should

[1] Ma Pan (1996-), Male, Gansu Qingyang, Master of Science, Capital Market and Governance of Listed Companies, Lanzhou University of Technology, China; Email: 1352353469@qq.com.

excess cash be held in order to fuel the innovation activities of manufacturing companies? Does excess cash holdings facilitate or inhibit innovation? This is a question that needs to be examined in depth.

The majority view is that excess cash holdings means that a company actually holds more cash than it needs to maintain its normal business activities. In recent years there has been a general tendency for listed companies in China to hold high levels of cash reserves. Research on the relationship between cash holdings and corporate innovation has received widespread attention from domestic academics: one view is that excess cash holdings have a crowding-out effect on corporate innovation performance due to the high opportunity cost [5]. Another view is that excess cash holdings have a catalytic effect on firms' innovation performance, and that the use of internal funds can effectively alleviate financing constraints to improve corporate innovation [6]. In conclusion, there is still disagreement on the relationship between corporate cash holdings and corporate innovation, and there is still room for further exploration.

To synthesise the above analysis, this paper empirically examines the relationship between corporate excess cash holdings and corporate innovation as well as the moderating role of product market competition in enhancing corporate innovation, using a sample of Chinese manufacturing companies listed on the Shanghai and Shenzhen A-shares from 2011-2021. The findings of the study have important implications for the promotion of corporate innovation.

2. Theoretical Analysis and Research Hypothesis

2.1 Excess cash holdings and corporate innovation in manufacturing firms

On the one hand, firms choose to set aside some excess cash within the firm based on speculative motives. As there is an additional opportunity cost associated with holding excess cash assets, firms tend to invest the excess cash to generate income to compensate for this opportunity cost. However, due to the imperfection of the capital markets and the lack of capital market savvy in manufacturing companies, manufacturing companies are more inclined towards corporate innovation than risky investments in the capital markets. On the other hand, innovation activities are long-term and require significant financial support, which can be obtained from internal and external sources. In reality, adequate cash reserves are the basis for innovation, and excess cash holdings can alleviate financing constraints and ensure continued investment in innovation activities. Hu Jun et al[7] conducted an empirical study based on the data of listed companies from 2007 to 2016 and found that long-term stable cash holdings play a crucial role in the innovation activities of high-investment firms. Dai Zhenzhen et al [8] took electronic equipment manufacturing enterprises as a sample and found through empirical evidence that electronic equipment manufacturing enterprises tend to use the excess cash holdings for corporate innovation investment thus improving their competitive position. Based on the above analysis, this paper proposes the following hypotheses:

H1: Manufacturing firms' excess cash holdings have a catalytic effect on corporate innovation.

2.2 Excess cash holdings, product market competition and firm innovation

The study of the relationship between excess cash holdings and firm innovation cannot avoid the competitive product market environment in which firms operate. On the one hand, product market competition has an information effect, and the transparency of corporate information is enhanced when competition is fierce, so that corporate owners and external investors have an invisible monitoring effect on agents through the comparison of transparent information. This enables agents to regulate their own behaviour and to deepen their understanding of the firm's objectives, so that they can use the firm's excess cash resources to invest in innovations that are more beneficial to the firm's future development, thus strengthening the role of excess cash holdings in promoting innovation. On the other hand, according to market process theory, when external market competition is intense, the original market equilibrium will be broken, thus making the product market move towards a higher equilibrium. Corporate profits. In a competitive market environment, the only way for firms to gain competitive advantage is to increase their investment in innovation. When firms face severe financing constraints, the risk of uncertainty is even greater. To avoid the risk of predation and obsolescence, firms often tend to hold excess cash to cope with possible market changes. Under the effect of product market competition, the relationship between excess cash holdings and the promotion of firm innovation will be further strengthened. Based on the above analysis, this paper proposes the following hypothesis:
H2: Other things being equal, product market competition positively moderates the relationship between a firm's excess cash holdings and firm innovation.

3. Research design

3.1 Sample selection and data sources

The initial research sample of this paper is the A-share listed companies in the manufacturing industry in Shanghai and Shenzhen in China during the period from 2011 to 2021, and all sample data are obtained from the Guotaian database. The sample was processed as follows: (1) Excluding the sample companies in the financial and insurance categories, ST and PT. (2) The continuous variables were subjected to an upper and lower 1% tailing process. The processing software used in this paper is Stata16 and Excel.

3.2 Description of variables

3.2.1 Explanatory variables.

The explanatory variable in this paper is corporate innovation (Rd), which is measured by the ratio of R&D investment to current operating income.

3.2.2 Explanatory variables

(i) Excess cash holdings (Excash)

Excess cash holding is the core explanatory variable of this paper. In the selection of variables, we refer to the method of Oplear et al [9] and use model 1 to construct the expected model of enterprise cash holding. The absolute value of the residuals from the model regression is used as a proxy variable for excess cash holdings, with a larger indicator indicating a higher level of excess cash holdings. The larger the indicator, the higher the level of excess cash holdings. The closer the indicator is to zero, the closer the firm's excess cash holdings are to normal levels.

$$\text{Lncash} = \beta_0 + \beta_1 Size + \beta_2 Cf + \beta_3 Nwc + \beta_4 Growth + \beta_5 Capex + \beta_6 Lev + \beta_7 Div + \sum Year + \varepsilon \quad (1)$$

(ii) Product market competition (Pcm)

For the measurement of the degree of product market competition, this paper mainly uses the Lerner index to measure the monopoly position of enterprises in the market, the Lerner index is calculated by the formula (price - marginal cost)/price. The Lerner Index is calculated as (price - marginal cost)/price. Because marginal cost itself is difficult to measure, this paper uses the profit margin instead, where Pcm = (operating revenue - operating costs - selling expenses - administrative expenses)/operating revenue. Its value varies between 0-1, the smaller the value of Pcm, the smaller the market power in the industry, the greater the competition faced by enterprises.

3.2.3 Control variables

As the relationship between excess cash holdings and corporate innovation can be affected by many factors, this paper refers to [2] Yu Yike et al. study in the selection of control variables to select corporate return on assets (Roa), Tobin's q, company size, total asset turnover, and the number of years the company has been listed as control variables. Detailed variable definitions and calculations are shown in Table 1.

Table 1 Definition of variables

Variable name	Variable Codes	Variable definition and calculation methods
Corporate Innovation	Rd	R&D investment / Current operating revenue
Excess cash holdings	Excash	Absolute value of residuals from regression of model (1)
Product market competition	Pcm	Lerner Index, pcm = (operating revenue - operating costs - selling expenses - administrative expenses) / operating revenue
Size of business	Lnsize	Natural logarithm of the company's total assets at the end of the year
Number of years the company has been listed	Age	Year of observation minus the natural logarithm of the year of incorporation plus 1
Return on assets	Roa	Return on Assets = Net Profit / Total Assets
Tobin q	TobinQ	Market value of business/total assets
Total asset turnover ratio	Turn	Operating income/total assets
Annual dummy variables	Year	Takes a value of 1 when the business is in that year, 0 otherwise

3.3 Model construction

In order to verify the relationship between excess cash holdings and firm innovation, this paper draws on the approach of Yu Desheng and Li Xing[10] to explore product financialization and firm innovation, and constructs model (2) as follows: $Rd_{i,t} = \beta_0 + \beta_1 Excash_{i,t} + \beta_2 Controls_{i,t} + \sum Year + \varepsilon_{i,t}$

To verify the moderating effect of product market competition on the relationship between excess cash holdings and firm innovation, other things being equal, this paper follows the method used by Desheng Yu and Xing Li[10] to verify the moderating effect and constructs model (3) as follows: $Rd_{i,t} = \gamma_0 + \gamma_1 Excash_{i,t} + \gamma_2 Pcm_{i,t} + \gamma_3 Excash_{i,t} * Pcm_{i,t} + \gamma_4 Controls_{i,t} + \sum Year + \varepsilon_{i,t}$
where $Excash_{i,t}$*$Pcm_{i,t}$ denotes the cross product term and Controls denotes a set of control variables listed previously; ε is the stochastic perturbation term of the model.

4. Empirical Results and Analysis

4.1 Descriptive statistics

Table 2 Descriptive statistics of the variables

Variables	Obs	Mean	Std. Dev.	Min	Max
Rd	16945	4.613	3.56	0.07	21.48
Excash	16945	0.167	0.078	0.011	0.358
Pcm	16945	0.138	0.104	-0.113	0.489
TobinQ	16945	2.081	1.227	0.882	7.998
Turn	16945	0.628	0.34	0.136	2.132
Roa	16945	0.052	0.048	-0.099	0.205
Lnisze	16945	22.02	1.163	20.025	25.595
Age	16945	2.833	0.345	1.74	3.482

From the descriptive statistics of the variables, it can be seen that the average value of enterprise innovation (the ratio of R&D investment to current operating income) is 4.613, which indicates that the innovation intensity of manufacturing enterprises in China is still very weak, with the minimum value of 0.07 and the maximum value of 21.48. It can be seen that there is a great difference in the innovation ability of different enterprises, which needs to be further strengthened. The maximum and minimum values of excess cash holdings are 0.011 and 0.358 respectively, indicating that the excess cash holdings of enterprises differ significantly among different manufacturing enterprises.

4.2 Regressivity analysis

Table 3 Regression results

	(Model 2)	(Model 3)
Variables	Rd	Rd
Excash	2.339*** (3.43)	4.890*** (5.45)
Pcm		0.339 (0.34)
Excashpcm		-17.077*** (-4.19)

TobinQ	0.058** (2.30)	0.063** (2.51)
Turn	-1.985*** (-12.73)	-2.108*** (-13.31)
Roa	-7.026*** (-10.01)	-4.145*** (-4.59)
Lnsize	-0.181* (-1.83)	-0.117 (-1.14)
Age	-0.291 (-0.63)	-0.302 (-0.66)
Year	Yes	Yes
Constant	9.353*** (3.80)	7.967*** (3.20)
Observations	16,945	16,945
R-squared	0.110	0.121
Number of name	2,658	2,658
adj_R2	0.120	0.120
F	23.37	23.37

Note: ① denotes 10% level of significance, ** denotes 5% level of significance, *** denotes 1% level of significance; ② t-statistics after Robust correction are in parentheses.*

Table 3 reports the results of the panel regressions on firms' excess cash holdings, product market competition and firm innovation. The second column indicates the relationship between excess cash holdings and corporate innovation, where the correlation coefficient of excess cash holdings is 2.339, which is significant at the 1% level, indicating that excess cash holdings of manufacturing firms listed in China have a significant role in promoting corporate innovation, verifying hypothesis 1. Excess cash holdings of manufacturing firms reduce the financing constraint in innovation investment, thus allowing firms to have sufficient funds for longer-cycle innovation R&D activities, resulting in the strengthening of the innovation capability of manufacturing firms.

The third column of Table 3 reflects the moderating effect of product market competition in the relationship between excess cash holdings and firm innovation, and the results show that the coefficient of the cross product term is -17.007, which is significant at the 1% level. In other words, the less competitive the product market is, the weaker the positive effect of excess cash holdings on firm innovation, and conversely, the weaker the inhibitory effect on the positive relationship between the two. The hypothesis of H2 is verified. According to the information effect of product market competition, an increase in product market competition makes firms disclose more information to the outside world, which is equivalent to imposing an "invisible oversight" on executives, which on the one hand alleviates agency conflicts and on the other hand avoids short-sightedness.

5. Conclusions and Recommendations

5.1 Research findings

This paper explores the impact of excess cash holdings on corporate innovation based on the perspective of product market competition, using a sample of manufacturing companies listed in Shanghai and Shenzhen A-shares from 2011 to 2021, and draws the following main conclusions: (1) Excess cash holdings of manufacturing enterprises have a catalytic effect on corporate innovation. The long cycle of corporate innovation activities and the high capital requirements of enterprises can effectively alleviate the financing constraints of enterprises and provide continuous financial support for their innovation activities. (2) Product market competition moderates the relationship between excess cash holdings and corporate innovation. The more competitive the product market is, the more significant the positive impact of excess cash holdings on corporate innovation, which is conducive to enhancing corporate innovation. The positive effects of the governance and predation effects of product market competition dominate. In terms of the governance effect, the more competitive the product market is, the better the external governance mechanism of the firm, which can curb the short-sighted and self-interested behaviour of executives and thus increase the investment in corporate innovation. In terms of the predatory effect, the more competitive the product market is, the more companies will take the initiative to increase their investment in innovation in order to prevent themselves from being "predated" by other companies in the same industry.

5.2 Countermeasures and suggestions

Based on the findings of this paper, the following suggestions are made: (1) Manufacturing enterprises can appropriately hold more funds than they need for normal operation according to their own situation, so as to alleviate the financing constraints in their innovation activities and reserve good funds for their innovation activities. (2) Enterprises should reasonably plan to retain corporate cash flow and formulate reasonable innovation strategies according to the degree of product market competition in the industry they operate in.

Reference

[1] Arrow K J. Economic Welfare and the Allocation of Resources for Invention[J]. social science electronic publishing, 1962.

[2]Kamien M I,Schwartz N L. Self-financing of an R and D project[J]. Am. Econ. Rev.; (United States), 1978, 68:3(3):252-261.

[3]Tang Qingquan,Xu Xin. Corporate R&D investment and internal funds--a study from Chinese listed companies[J]. China Accounting Review,2010,8(03):341-362.

[4] Lu X., Zheng Y. F., Li J. M. A study on the impact of financing constraints on corporate R&D investment--empirical evidence from high-tech listed companies in China[J]. Accounting Research,2013(05):51-58+96.[5] Lydia Guo. The impact of financialization of manufacturing on innovation investment: "crowding out effect" or "reservoir effect"? [J]. Modern Economic Inquiry,2017(12):49-59.

[6] Yu Yike, Guo Jing, Zhang Zaijie et al. Real estate crowding, cash holdings and firm innovation performance[J]. Statistics and Decision Making,2020,36(03):159-163.

[7]Hu J,Li Q,Zeng Y. The smoothing effect of cash holdings on R&D investment:A product market competition perspective[J]. Journal of Management Engineering,2021,35(06):150-162.

[8] DAI Zhenzhen,CHENG Kequn. Cash holdings, R&D investment and firm competitiveness[J]. Journal of Guangxi University of Science and Technology,2019,30(04):118-126.

[9] Opler T, Pinkowitz L, Stulz R, et al. The determinants and implications of corporate cash holdings[J]. Journal of financial economics, 1999, 52(1): 3-46.

[10] Yu, Desheng, Li, Xing. Corporate financialization, product market competition and corporate innovation[J]. Financial Development Research,2021,No.471(03):20-25.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230180

Research on the Evaluation of Economic Responsibility Audit of State-Owned Enterprise Leaders from the Perspective of Green and Low Carbon

Jing LIU[①], Li MA and Cuilin WANG
School of Economics and Management, Lanzhou University of Technology, No. 287, Langongping Road, Qilihe District, Lanzhou City, Gansu Province, China

Abstract. With the implementation of the "carbon peaking and carbon neutrality" target strategy, green, low-carbon development has become the main way of economic development, and the low-carbon economic operation and healthy development of state-owned enterprises not only play a leading role in the operation of the national economy as a whole, but also are economic entities that bear greater pressure for green and low-carbon development. This paper focuses on the construction of an evaluation system for the economic responsibility audit of the leaders of state-owned enterprises from a green and low-carbon perspective, using the hierarchical analysis method to determine the weights of each evaluation index, and selecting three commercial state-owned enterprises of Gansu province for verification. The evaluation system is feasible and contributes to the balanced development between the economic development of state-owned enterprises and green and low-carbon.

Keywords. green and low carbon; state-owned enterprises; economic responsibility audit; audit evaluation system

1. Introduction

Since the 1990s, China's economic responsibility audits have made great strides in development, with the successive introduction of relevant systems and regulations. The "Regulations on Economic Responsibility Audits of Major Leading Cadres of the Party and Government and Major Leaders of State-owned Enterprises and Institutions", revised in 2019, state that economic responsibility audits should implement the new development concept, promote high-quality economic development, facilitate comprehensive deepening of reform and promote modernization of the national governance system and governance capacity. In the report of the 20th Party Congress, it was stressed that the concept of "green water and green mountains are the silver mountain of gold" must be firmly established and practiced. Against the backdrop of the "double

---

① Corresponding Author, Jing Liu, School of Economics and Management, Lanzhou University of Technology, Lanzhou, Gansu, China. Email: 1764081643@qq.com.
Fund project: Gansu Provincial Audit Department key scientific research project "State-owned enterprises economic responsibility audit evaluation system construction and application research - a commercial class of state-owned enterprises in Gansu Province as an example" (SJ210401) project funding

carbon" goal of achieving peak carbon and carbon neutrality, strengthening ecological environmental protection and actively responding to climate change with a view to reducing pollution and reducing carbon as a means to increase efficiency is the only way to build a "beautiful China" and achieve sustainable development for the Chinese nation Li and Fang（2023)[1] . State-owned and state-controlled enterprises are the "front-runners" in social and economic development, and are an important force in developing a green economy and building an ecological civilization Liang L. (2021)[2], Li Chao(2021)[3]. Therefore, proposes to reconstruct the evaluation system based on the existing evaluation system for the economic responsibility of leading cadres of state-owned enterprises, with emphasis on the need for full coverage and green and low-carbon development, using a commercial class of state-owned enterprises as an example and three entities in Gansu Province as a trial run.

2. Construction of an evaluation system for economic responsibility audits of state-owned enterprises in the commercial category

2.1. Research status of evaluation index construction

In the process of constructing the evaluation index of enterprise performance audit, different scholars adopt different methods, but they always follow the analysis principle of qualitative and quantitative merger. For example :Zheng Shiqiao (2018)[4] ,Wang Xiaohui（2006)[5],Li ZD, Guo L.（2022)[6]believes that the economic responsibility evaluation index system should be based on different industry characteristics and different types of cadres to select economic responsibility evaluation indicators. In the context of low-carbon economic development, Liu Xuan（2015[7]Zheng Guohong (2017)[8]Chen Yijin(2023)[9] used AHP to construct a comprehensive audit evaluation system for green economic responsibility of state-owned enterprises, mainly selecting indicators from three levels : sustainable development, production and operation, and social responsibility.

2.2.Construction of the evaluation system of commercial state-owned enterprises

The evaluation system for commercial state-owned enterprises is ② guided by the concepts of full audit coverage, "carbon peaking and carbon neutrality" strategy and green low-carbon development. Specifically, it includes: the guideline level, i.e. the primary indicators: return on capital and quality efficiency, marketability, internationalization, sustainability of economic development, and integrity and governance. Specifically, there are 5 first-level indicators, 13 second-level indicators and 47 third-level indicators.(see Table 3 for details).

② Based on the role, status quo and needs of SOEs in economic and social development, central enterprises are defined as commercial (including commercial category 1 and commercial category 2) and public welfare according to their main business and core business scope, including: commercial category 1 enterprises: focus on fully competitive industries and fields, improve the rate of return on state-owned assets, ensure product quality, and be a market-oriented, international and competitive leader. The enterprises in the commercial category are

3. Calculation and application of the weights of the economic responsibility audit evaluation indicators

3.1.Determination of the weighting of audit evaluation indicators

This paper applies the hierarchical analysis method (AHP) to determine the weights of economic responsibility audit evaluation indicators for a commercial category of state-owned enterprises.

3.1.1 Modelling the hierarchy. Objective level: Evaluation index system for economic responsibility audits of commercial category 1 state-owned enterprises.

Criteria level: The five dimensions of return on capital and quality and efficiency, marketability, internationalization, sustainability of economic development and integrity are set out in 13 levels, all of which are positive indicators, B1, B2, B3,...,B13 respectively. Indicator layer: 47 items in total, of which 42 are positive indicators, 5 are negative indicators and 3 are qualitative indicators. These are C1,C2,C3,...,C47. See Table 2:

3.1.2 Construction of the judgment matrix.

The construction of the judgment matrix starts from the 2nd layer of the hierarchical model, and the indicators in the same layer are compared two by two from top to bottom layer by layer to obtain the judgment matrix B = $(b_{ij})n*n$, b_{ij} denotes the relative weight values of elements i and j in a layer and the elements in the previous layer, i.e. the values obtained by comparing the two elements of i and j. n denotes the number of elements, then the calculation formula of the judgment matrix A is as follows: B = $(b_{ij})n*n$
The matrix is characterized as follows: $b_{ij} > 0$ $b_{ij} = 1/b_{ij}$ $b_{ij} = 1$
Taking the profitability of the elements of the secondary indicator B1 as an example, by constructing a judgment matrix of the above elements that are comparable between two elements and using the metrics to give the corresponding weights to each element, such as the judgment matrix constructed for the B1 criterion level and the indicator level below B1 is shown in Table 1[③] :

Table 1. B1 Judgement matrix constructed at the criterion level and the corresponding indicator level

B1	C1	C2	C3	C4	C5	Wi
C1	1.0000	0.2500	0.3333	0.2500	0.2000	0.0554
C2	4.0000	1.0000	2.0000	1.0000	0.3333	0.2018
C3	3.0000	0.5000	1.0000	1.0000	0.3333	0.1444
C4	4.0000	1.0000	1.0000	1.0000	0.5000	0.1905
C5	5.0000	3.0000	3.0000	2.0000	1.0000	0.4079

3.1.3.Consistency test and calculation of criterion layer weights.

The following is an example of the process of calculating the weights for the profitability profile B1 at the return on capital and quality return levels:① The product of the values

[③] The data in the table indicates which of the two elements below B1, Ci or Cj, is more important and influential, on a scale of 1 to 9. The degree of influence bij, which indicates Ci and Cj on B1, is quantified by assigning values.

of the elements of each row of matrix A yields M_i . i.e. calculate the geometric mean of the factors in each row $M =_i \prod a_{ij}$, i = 1, 2, ..., n, to obtain $M_1 = 0.0042$, $M_2 = 0.6666$, $M_3 = 0.1667$, $M_4 = 0.5$ and $M_5 = 18$.②Calculate the nth root of M_i to obtain $W =_i \sqrt[n]{Mi}$. Calculate the 5th root of M_i , i.e. $Wi=\sqrt[n]{Mi},$ i = 1, 2, ..., n, to obtain $W_1 = 0.3342$,$W_2 = 0.9221$,$W_3 = 0.6988$,$W_4 = 0.8706$,$W_5 = 1.7826$.③Normalize the vector to obtain $W_i = M /_i \sum_{j=1}^{n} Mj$. W1 = 0.0554, W2 = 0.2018, W3 = 0.1444, W4 = 0.1905 and W5 = 0.4079 were calculated, i.e. the resulting eigenvector is W = (0.0554, 0.2018, 0.1444, 0.1905, 0.4079).④ Calculate the maximum eigenvalue $\lambda\ max = \frac{1}{n}\sum_{i=1}^{n}\frac{(B_1 w)_i}{w_I}$. This yields $\lambda\ max = 5.1160$.⑤Consistency test, using the random consistency ratio CR to determine whether the matrix has a head and tail consistency, the test standard is : CR = CI / RI. If CR ≥ 0.1, the judgment matrix is chaotic and lacks consistency, and the elements of the judgment matrix need to be adjusted. If CR < 0.1, the judgment matrix is consistent. By calculating the CR values of each index are less than 0.1, as shown in the table 2. All pass the consistency test, and the obtained weight results are reasonable and can be used.

Table 2.CR value of each evaluation index

Indicators	Target level	(B1)	(B2)	(B3)	(B4)	(B5)	(B6)
CR	0.0517	0.029	0.0629	0.0092	0.0274	0.0274	0.0001
Indicators	(B7)	(B8)	(B9)	(B10)	(B11)	(B12)	(B13)
CR	0.0437	0.0001	0.0001	0.0001	0.0001	0.0239	0.0072

Based on the above calculation steps, the AHP software was used to calculate the weights of each criterion layer and the corresponding indicator layer one by one. See Table 3:

Table 3. Evaluation index system and weights for economic responsibility audits

Target level	Guideline level				Indicator layer	Weighting
	Tier 1 indicators		Secondary indicators		Tertiary indicators	
Economic responsibility audit evaluation index system for commercial category 1 state-owned enterprises 100%	Return on capital and quality gains	24%	Profitability (B1)	9.88 %	Return on Net Assets ($C_{1,1}$)	0.55%
					Return on Total Assets ($C_{1,2}$)	1.99%
					Surplus cash cover multiple ($C_{1,3}$)	1.43%
					Cost Margin ($C_{1,4}$)	1.88%
					Rate of return on capital ($C_{1,5}$)	4.03%
			Asset quality (B2)	7.91 %	Total Asset Turnover ($C_{2,1}$)	1.82%
					Inventory turnover rate ($C_{2,2}$)	0.37%
					Accounts receivable turnover rate ($C_{2,3}$)	2.17%
					Non-performing assets ratio ($C_{2,4}$)	1.15%
					Cash recovery rate on assets ($C_{2,5}$)	2.40%
			Debt risk (B3)	4.31 %	Gearing ratio ($C_{3,1}$)	0.92%
					Interest earned multiplier ($C_{3,2}$)	0.40%
					Current ratio ($C_{3,3}$)	0.25%

					Quick Ratio ($C_{3,4}$)	0.85%
					Cash flow liability ratio ($C_{3,5}$)	1.89%
			Operational growth (B4)	1.90 %	Sales (operating) growth rate ($C_{4,1}$)	0.09%
					Capital preservation and appreciation rate ($C_{4,2}$)	0.44%
					Sales (operating) profit growth rate ($C_{4,3}$)	0.72%
					Growth rate of total assets ($C_{4,4}$)	0.42%
					Technology input ratio ($C_{4,5}$)	0.23%
	Level of marketability	11.12%	Degree of market openness (B5)	4.69 %	Market-based employment levels (qualitative) ($C_{5,1}$)	0.48%
					Employee shareholding ratio ($C_{5,2}$)	0.81%
					Amount of dividends ($C_{5,3}$)	3.40%
			Degree of market development (B6)	2.73 %	Extent to which an enterprise's internal market-based management system is well developed (qualitative) ($C_{6,1}$)	1.36%
					The degree of adequacy of internal supporting departmental regulations (qualitative) ($C_{6,2}$)	1.36%
			Market share (B7)	3.70 %	Overall market share ($C_{7,1}$)	0.22%
					Target market share ($C_{7,2}$)	0.69%
					Relative market share ($C_{7,3}$)	1.02%
					Comparative market share ($C_{7,4}$)	1.77%
	Internationalization level	17.42%	Inward-looking internationalisation (B8)	3.55 %	Ratio of foreign purchases to total purchases ($C_{8,1}$)	1.18%
					Ratio of foreign investment attracted to total investment ($C_{8,2}$)	2.37%
			Outward-looking internationalisation (B9)	6.09 %	Ratio of foreign investment to total investment ($C_{9,1}$)	2.03%
					Foreign sales to total sales ratio ($C_{9,2}$)	4.06%
			Invisible Internationalisation (B10)	7.78 %	"One Belt, One Road" exchange ($C_{10,1}$)	2.59%
					Outbound investment profile ($C_{10,2}$)	5.19%
	Economic development sustainability	36.41%	Social contribution (B11)	17.59 %	Social contribution rate ($C_{11,1}$)	5.86%
					Employment contribution rate ($C_{11,2}$)	11.73 %
			Environmental	18.82 %	Carbon emission rate ($C_{12,1}$)	2.74%

			protection and resource use (B12)		Ecosystem restoration rate ($C_{12,2}$)	1.82%
					Rate of treatment of "three wastes" ($C_{12,3}$)	1.79%
					Environmental profit and loss lift rate ($C_{12,4}$)	1.86%
					Rate of reduction in energy consumption per unit of GDP ($C_{12,5}$)	3.53%
					Rate of reduction in resource consumption ($C_{12,6}$)	3.45%
					Rate of conversion of scientific and technological achievements ($C_{12,7}$)	3.63%
	Integrity in Politics	11.05%	Integrity in Politics (B13)	11.05%	Penalty expenditure rate ($C_{13,1}$)	1.30%
					Personal Illicit Income Rate ($C_{13,2}$)	3.51%
					Average annual duty consumption ratio ($C_{13,3}$)	6.24%

3.2. Application of evaluation indicator system

3.2.1. Data collection.

Quantitative indicators data were obtained from Gansu Province A, B and C[4] Financial data of three commercial category 1 listed state-owned enterprises for 2019-2021 were compiled and summarized by hand. Qualitative indicators are transformed into quantitative data through the standardization of data, based on the ideas and application of the fuzzy mathematical affiliation theory of L.A. Zadeh (1965), Professor of Cybernetics at the University of California, USA.

3.2.2. Evaluation Indicator Normalization Processing.

The raw data values were normalized to keep the results within [0, 1]. Firstly, a distinction is made between very large and very small indicators of their own nature, and secondly, the indicators are calculated based on the formula:

Very Large Indicators: $x_{ij} = \frac{u_{ij}}{uj_{\max}}$ Very Small Indicators: $x_{ij} = \frac{uj^{\min}}{u_{ij}}$

[4] As it relates to the financial information of the business, A, B and C are used instead of the company name.

3.2.3.Hierarchical analysis Evaluation.

According to the classification of target level, criterion level (including primary and secondary indicators), indicator level (including tertiary indicators) and the calculated weights in Table 2, and combined with the collected data of specific indicators, the state values of each level Z, B_i , C_{ij} are calculated by the formula: $Z= \sum \text{Factor status value} \times weighted$.First, the score of the criterion level (secondary indicator) is calculated, $B_i=$ $\sum \text{Level 3 indicator status values Cij}$⑤ $\times \text{Three} - \text{level indicator weights Wi}$, and the calculation results are shown in Table 3 for B1..., B13.Next, the target layer score was calculated, $Z= \sum \text{Bi} \times \text{Benchmark layer secondary indicator weights di}$, and the calculation results are shown in Table 4 for Z values.

Table 4. Table of comprehensive evaluation results

Company A	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12	B13	Z
	0.042	0.008	0.210	-1.972	0.009	0.512	0.205	0.139	0.108	0.597	-0.009	0.337	0.000	0.082
Enterprise B	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12	B13	Z
	0.056	0.115	0.148	0.469	0.001	0.270	0.079	0.041	0.038	0.360	0.008	0.000	0.000	0.073
Enterprise C	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12	B13	Z
	0.062	0.116	0.264	0.205	0.003	0.544	0.414	0.101	0.084	0.363	0.009	0.010	0.000	0.100

3.2.4.Evaluation and analysis of the results.

According to the evaluation and analysis results of the three enterprises, the evaluation values of enterprises A, B and C are 0.082, 0.073 and 0.100 respectively, with C > A > B, indicating that enterprise C has the best performance in fulfilling its economic responsibility, followed by enterprise A and the worst by enterprise B. From the primary indicators, it is clear that enterprises A, B and C are all performing well in terms of their economic responsibilities in terms of integrity and meet the compliance objectives of the economic responsibility audit.

A more specific analysis reveals that Enterprise C ranks first among the three enterprises in terms of economic responsibility performance in the three areas of return on capital and quality of earnings, marketability and social contribution at the level of primary indicators. Looking specifically at the secondary indicators, it is found that the performance of economic responsibility fulfillment needs to be further improved in the areas of operational growth, market openness, internationalization, environmental protection and resource utilization. Company A is the strongest in terms of internationalization, environmental protection and resource utilization, but the weakest in terms of social contribution. Company C is the best overall, particularly in terms of quality of operations, market openness, internationalization and social contribution, but is clearly weaker in terms of environmental protection and resource utilization.

⑤ Collecting and researching data

4.Conclusions and recommendations

The above study shows that focusing on the green and low-carbon dimension in the economic responsibility audit of leading cadres of state-owned enterprises will definitely help to enhance the environmental awareness of enterprises and balance between enterprise development and low carbon, thus helping to achieve the "double carbon" goal of China rapidly. We suggest: Firstly, green, low-carbon development requires a lot of investment upfront, and in order to sort out good low-carbon development concepts in enterprises, governments at all levels need to make more development policies and financial investments to encourage this, while strengthening economic responsibility audit evaluation. Second, the government should guide enterprises to establish low-carbon development funds, especially for enterprises with good economic development, and encourage a good balance between economic development and low carbon.

References

[1] Li P,Fang J. Environmental regulation, digital economy and corporate green innovation. Statistics and Decision Making. 2023 Mar: 39(05): 160-165.

[2] Liang L. Research on innovation of enterprise economic responsibility audit model under ESG framework [J]. Finance and Accounting,2021 Mar; 630(06): 52-54.

[3] Li Chao. On the construction of evaluation index system of internal economic responsibility audit in agricultural scientific research institutions [J]. Agricultural economy, 2021, No.406 (02) : 36-37.

[4] Zheng Shiqiao. Evaluation of economic responsibility audit of leading cadres : theoretical framework and example analysis [J]. Accounting Monthly, 2018 (21) : 139-145

[5] Wang Xiaohui. The construction of three-dimensional model mechanism of economic responsibility audit of state-owned enterprises [J]. Audit and economic research, 2006 (03) : 44-47

[6]Li ZD, Guo L.The evaluation measurement of economic responsibility audit under high quality development--Jiangsu Province as an example[J]. Finance and Accounting Monthly, 2022 Jan; 921(05): 79-85.

[7] Theoretical Research Group of Jinan Special Office of Audit Commission. Liu Xuan, Wang Guian. Audit Research on State-owned Enterprises under the Background of Comprehensively Deepening Reform [J]. Audit Research, 2015 (02) : 36-41

[8] Zheng Guohong. AHP model thinking and improvement of green economic responsibility audit evaluation of state-owned enterprises [J]. Journal of Southwest University of Political Science and Law. 2017 (01) : 97-104

[9] Construction of performance evaluation system for high-quality development of Chongqing state-owned enterprises under the ' double carbon goal ' [J]. Chinese and foreign corporate culture, 2023, No.639 (02) : 114-116.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230181

Research on Location-Route Problem of New Energy Logistics Vehicle Distribution System

Hao CHEN[a] and Xuerui Qin[a1]
[a] *School of Economics and Business Administration, Yibin University, Yibin City, Sichuan, China.*

Abstract. This paper takes the comprehensive system problem of charging station location and shortest path optimization faced by new energy vehicles in the distribution process as the research object. We hope to obtain a combination optimization method that can further support China's use of new energy vehicles to complete urban distribution problems. This paper addresses the existing comprehensive challenges related to route selection and site allocation. It takes into account the short transportation mileage and low load capacity of new energy vehicles caused by charging constraints. At the same time, considering the timeliness of modern urban distribution, time window constraints and logic constraints are added. Taking the minimum number of charging stations and the minimum transportation cost as the multi-objective problem of combinatorial optimization, the mathematical model is finally established. After transforming the multi-objective problem into a single-objective problem, the genetic algorithm the genetic algorithm is used to solve the system problem comprehensively. Finally, an example is used to verify the feasibility and effectiveness of genetic algorithm in solving the new energy vehicle distribution - location path combination optimization problem. It provides theoretical support and development prospects for further promoting urban distribution of new energy vehicles in the future.

Keywords. New energy vehicles, System location, routing problem, genetic algorithm

1. Introduction

In recent years, the rapid development of China's economy, the rapid advancement of urbanization, and the huge transformation of consumption patterns have brought new challenges to urban logistics and distribution. New energy vehicles are the most promising means of transportation for future urban logistics distribution, with many advantages such as environmental protection, low-carbon, and low noise. However, due to the unique performance characteristics of new energy vehicles, the application of new energy logistics vehicles in distribution cannot simply draw on the operating experience of traditional fuel trucks. For new energy logistics vehicles, their logistics management, driving environment, energy replenishment, and mileage limitations are all new issues

[1] Corresponding Author, Xuerui Qin, Yibin University, Yibin City, Sichuan, China; Email: 2019114013@yibinu.edu.cn ; Funding Support: Yangtze River Upstream Shipping and Logistics Collaborative Innovation Center (XTCX2022A02).

faced in distribution. How to establish and improve the logistics network distribution system for new energy vehicles, the selection of supporting facilities, and the optimization of distribution routes have become important issues that need to be urgently solved.

In terms of path optimization problems, Erdogan and Miller-hook first considered the charging station factor in the modeling of green vehicle path problems, with the overall goal of minimizing the total delivery mileage [1-2]. Chen proposed a time scheduling model that considers the degradation effect of batteries and achieves orderly charging of vehicles according to the time schedule [3]. Fu Zhengtan developed an electric vehicle path optimization problem that considers time windows, and took the lead in considering charging behavior on the road [4]. DOMINIK GOEKE has expanded this model, taking into account different types of hybrid electric logistics vehicles[5].

In terms of location problem, Wu Di. embedded transportation cost, time limit, warehouse operation cost and other parameters into the model analysis, modified the flow model of goods, and finally obtained the optimal location of the third-party logistics company's distribution center[6]. Mekhum found a logistics distribution center address that meets the logistics service needs of various major logistics service objects by conducting a sampling questionnaire survey on the location selection of logistics enterprise distribution centers[7]. Maximilian Schiffe takes the research on the location selection of distribution centers in St. John's Port as an example, introduces a virtual transportation logistics model, and obtains the optimal configuration address, effectively reducing the cost of agricultural product distribution for many small farmers[8].

This article takes into account the current popular situation of new energy vehicle delivery, and further adds constraints by considering the limited delivery mileage and strict load constraints of new energy vehicles. In order to solve the more complex location path combinatorial optimization problem in this paper, genetic algorithm is selected to solve the optimization problem in this paper by using its convergence speed block, which can jump out of the local optimal solution to the global optimal solution, which proves the feasibility of this method in solving the location path combination problem.

2. Design of New Energy Vehicle Delivery Site Selection and Path Optimization Model

2.1. Conditional Assumptions

It is necessary to fully consider the actual situation and the characteristics of new energy vehicles, such as weather conditions, congested road conditions, sudden situations, etc. Therefore, the following assumptions are made:

(1) All delivery vehicles depart from the same distribution center to complete the delivery task.

(2) During the delivery task, no new delivery tasks are accepted and the goods carried meet the demand at the beginning of the task.

(3) During transportation, weather, traffic congestion, and unexpected situations are not considered.

(4) For any delivery point, up to one vehicle can complete the delivery task.

(5) Considering various factors such as the optimal cruising range with a maximum speed of 120km/h and 80km/h, as well as the speed limit of 60km/h on urban roads, the

speed of new energy vehicles designed in this article is uniformly driven at a speed of 60km/h.

(6) Considering the appropriate load capacity and maximum transportation distance, in order to better complete the delivery task, the design limits the load capacity of a single vehicle to less than 3 tons, and the transportation distance is 160km.

2.2. Parameter Description

The parameter assumptions are shown in Table 1.

Table 1 Parameter Definition

parameter	meaning	parameter	meaning
p_0	Represents a distribution center point	LT_i	Indicates the latest time allowed for delivery vehicles to arrive at store i
p_i	Indicates the i-th customer demand point	t_{ij}^k	Indicates the transportation time from store i to j
M	Represents a set of customer numbers	RT_i	Indicates the time when the delivery vehicle arrived at store i
K	Represents the collection of vehicles in the distribution center	β_1	Unit penalty cost for early arrival
d_{oi}	Indicates the distance between the distribution center and store i	β_2	Unit penalty cost for late arrival
d_{ij}	Indicates the distance between store i and store j	a_1	Deterioration rate of goods during transportation
S_i	Indicates the demand for products in each store (in tons)	a_2	Deterioration rate of goods during unloading
C_k	Fixed cost of the k-th vehicle	t_{ij}^k	Transport time of the k-th vehicle from i to j
q_k	Indicates the loading capacity of the k-th delivery vehicle	U_j	The unloading time of the delivery truck at the store
α	Represents the transportation and distribution cost per unit distance	r_i	Mass of cargo compartment when leaving cargo i
C_v^k	Represents the driving cost incurred by the k-th vehicle	P_i	Is it necessary to set up a charging station at the i-th store? 1 is set, 0 is not set
ET_i	Indicates the earliest allowed delivery vehicle arrival time for store i		

Due to the time characteristics of customer demand involved in urban delivery, customers generally require goods to be delivered within a certain time interval, which must be within a time acceptable to customers in order to meet their requirements. If delivered too early or too late, dissatisfaction will occur and penalty costs will be incurred. This is a soft time window constraint, RT_i where the delivery vehicle arrives within the interval $[ET_i, LT_i]$; If customer point j is on the same delivery route CT_j , $CT_j = RT_i + UT + T_{ij} - RT_j$ then $RT_i = T_{oi}$.

To facilitate the establishment of the model, binary variables are defined as follows:

$$x_{ij}^k = \begin{cases} 1 & \text{Vehicle k travels from i to j} \\ 0 & \text{else} \end{cases} \tag{2-1}$$

$$y_{ij}^k = \begin{cases} 1 & \text{The demand for i is met by vehicle k} \\ 0 & \text{else} \end{cases} \tag{2-2}$$

2.3. Objective function

2.3.1 Total transportation cost

(1) Fixed costs: The fixed cost of vehicles, including expenses such as vehicle consumption, maintenance, and driver's wages:

$$C_1 = \sum_{k=1}^{K} C_k \tag{2-3}$$

(2) Transportation costs: Transportation costs refer to the electricity, repair, and rental costs incurred by new energy vehicles during transportation. These variable transportation costs are divided into transportation distances, so the transportation costs are as follows:

$$C_2 = \sum_{k=1}^{K}\sum_{i=1}^{M}\sum_{j=1}^{M} C_v^k d_{ij} x_{ij}^k \tag{2-4}$$

(3) Cost of damage to goods: The cost of goods loss refers to the loss of goods that occurs during the transportation and unloading process of the product. The cost of goods loss is the cost of transportation plus the cost of unloading:

$$C_3 = \alpha \sum_{k=1}^{K}\sum_{i=1}^{M}\sum_{j=1}^{M} x_{ij}^k \left(a_1 t_{ij}^k + a_2 U_j\right) r_i \tag{2-5}$$

(4) Time penalty cost: Due to the special nature of perishable fresh food, it is necessary to complete the delivery task within the customer's acceptable service time. If the food is not delivered to the demand point within the customer's acceptable time, there will be penalty costs.

$$C_4 = \beta_1 \sum_{i=1}^{M} \max\left(ET_i - RT_i, 0\right) + \beta_2 \sum_{i=1}^{M} \max\left(RT_i - LT_i, 0\right) \tag{2-6}$$

The total cost of delivery is:

$$\begin{aligned} \min C = C_1 + C_2 + C_3 + C_4 = \sum_{k=1}^{K} C_k + \sum_{k=1}^{K}\sum_{i=1}^{M}\sum_{j=1}^{M} C_v^k d_{ij} x_{ij}^k + \alpha \sum_{k=1}^{K}\sum_{i=1}^{M}\sum_{j=1}^{M} x_{ij}^k \left(a_1 t_{ij}^k + a_2 U_j\right) r_i \\ + \beta_1 \sum_{i=1}^{M} \max\left(ET_i - RT_i, 0\right) + \beta_2 \sum_{i=1}^{M} \max\left(RT_i - LT_i, 0\right) \end{aligned} \tag{2-7}$$

2.3.2 Number of charging stations

The number of charging stations set is appropriate. Having too few charging stations can cause vehicles to wait for charging, making it inconvenient to use and unable to support regional distribution tasks. However, having too many charging stations can result in significant cost and road pressure, especially in cities that are currently using new energy vehicle distribution services, which are mostly large cities or super large cities with high traffic pressure, such as first and second tier cities. Traffic is relatively congested and land space is difficult. Therefore, the design of the number and location of charging stations is also the goal of this study, but it should be designed to minimize the number of charging stations as much as possible after meeting the objective function (2-5). This is the most reasonable and convenient way for new energy vehicles of distribution enterprises to complete distribution tasks.

$$\min P = \sum_{i=1} P_i \tag{2-8}$$

2.4. Constraints

$$\sum_{k} y_{ki} = 1 \qquad i \in M \tag{2-9}$$

$$y_{ki} = \{0,1\} \quad i \in M; \forall A \tag{2-10}$$

$$x_{ki} = \{0,1\} \quad i \in M; \forall A \tag{2-11}$$

$$\beta_{ijk} > 0 \quad i \in M; \forall A \tag{2-12}$$

$$\sum_{i} x_{ijk} = y_{ik} \quad i \in M; \forall A \tag{2-13}$$

$$\sum_{j} x_{ijk} = y_{ik} \quad j \in M; \forall A \tag{2-14}$$

$$RT_i \in [50\% ET_i, 150\% LT_i] \quad i \in M \tag{2-15}$$

$$\sum_{k=1}^{K} q_k \leq 3 \tag{2-16}$$

$$\sum_{i=1}^{M} \sum_{j=1}^{M} d_{ij} x_{ij}^{k} \leq 100 \tag{2-17}$$

Meet the following constraints. Among them, equations (2-9) indicate that each demand point has only one vehicle for delivery; Equations (2-10) and (2-11) are integer constraints; Equation (2-12) refers to the unit time loss cost caused by the failure to deliver according to the customer's requested time; Equations (2-13) and (2-14) indicate that the delivery route of the delivery vehicle is a closed route; Equation (2-15) indicates that the product has arrived within the delivery time required by the customer; Equations (2-16) represent the allowable load constraints for new energy vehicles; Equation (2-17) represents the maximum mileage constraint for new energy vehicles.

3. Genetic Algorithm for Solving Element Design

The basic steps of genetic algorithm are:

(1) Encoding: Encoding is generally described in mathematical language in a certain order based on the final result of the problem. However, choosing different coding forms will have a great impact on the running time, running speed and fitness value of the algorithm, and even directly affect whether the optimal solution can be found in the end.

(2) Selection operator: In order to ensure that the genetic algorithm can jump out of the local optimal solution and approach the global optimal solution, it is necessary to set the genetic rules of the parents and children, namely the selection operator.

(3) Crossover operator: In the process of inheritance from parents to offspring, organisms recombine chromosomes through certain crossing rules to generate new offspring individuals. Only by continuously increasing the number of offspring individuals can the optimal solution be found.

(4) Mutation operator: Due to the crossover operator retaining a segment of dominant genes to the next generation, it is possible to cause offspring individuals to fall into local optima. Therefore, a segment of genes must be set for random mutation.

(5) Fitness value: The greater the fitness value, the better the individual is, the greater the probability of its genes being passed on to the next generation, and vice versa.

(6) Termination Rules: If the optimal solution is not found, it will constantly jump between several local optimal solutions and be difficult to converge. Considering the actual situation, it is necessary to design a termination rule, which often determines the total number of iterations.

4. Example verification

4.1. Background Description

Taking the data collected and organized by a certain enterprise in a city as an example, the enterprise currently has a distribution center (0 points) and 30 demand nodes (1-30 points). The specific coordinates and demand parameters are shown in Table 4-1. In addition, in urban areas, there are 10 coordinate points (C1-C10) where charging stations can be set up, and the coordinates of the 10 charging stations are shown in Table 4-2. The ultimate goal is to choose several routes starting from the distribution center point, around the remaining 50 coordinates and returning to point 0, with the shortest distance traveled; At the same time, it is also considered that in the proposed charging station C1-C10, the minimum number of constructions will be constructed[9].

Table 2 Distribution Center and Demand Node Coordinates and Demand Volume

Serial number	0	1	2	3	4	5	6	7	8	9	10
x	15	80	19	49	45	65	71	76	28	68	66
y	43	25	93	35	20	26	62	48	36	84	59
requirement	/	1140	990	840	930	870	660	750	690	720	750
Serial number	11	12	13	14	15	16	17	18	19	20	21
x	17	12	50	96	35	59	23	76	26	51	70
y	55	92	29	76	76	39	57	8	6	54	78
requirement	870	630	1170	1170	900	900	810	1170	840	690	1080
Serial number	22	23	24	25	26	27	28	29	30		
x	90	96	55	14	15	26	85	26	82		
y	94	13	57	47	2	34	17	80	32		
requirement	840	750	870	660	690	1170	1200	960	660		

Among them, the unit of coordinate distance is km, and the unit of demand is kg.

Table 3 Coordinates of the Proposed Charging Station

Serial number	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
x	94	68	76	75	40	66	18	71	4	28
y	5	10	83	70	32	96	4	44	39	77

Considering the daily use of the vehicle and battery loss, the maximum mileage of a single vehicle is 160km, and the maximum load capacity of a single vehicle is 3t. In order to transform the objective function into a single objective function, the minimum number of locations for charging stations is also converted into cost, that is, the construction cost of charging stations is 50, and the objective functions 3-5 and 3-6 are transformed into:

$$\begin{aligned}\min C = C_1 + C_2 + C_3 + C_4 = \sum_{k=1}^{K} C_k + \sum_{k=1}^{K}\sum_{i=1}^{M}\sum_{j=1}^{M} C_v^k d_{ij} x_{ij}^k + \alpha \sum_{k=1}^{K}\sum_{i=1}^{M}\sum_{j=1}^{M} x_{ij}^k \left(a_1 t_{ij}^k + a_2 U_j\right) r_i \\ + \beta_1 \sum_{i=1}^{M} \max\left(ET_i - RT_i, 0\right) + \beta_2 \sum_{i=1}^{M} \max\left(RT_i - LT_i, 0\right) + 50 \times \sum_{i=1} P_i\end{aligned} \quad (4\text{-}1)$$

Considering that the cost as the fitness value will lead to an increase in solving time and calculation, the calculation of its fitness value is the shortest total mileage in the process of using the genetic algorithm[10].

4.2. Output Results

The following is an introduction to the experimental results of genetic algorithm, without additional restrictions such as road conditions. The main steps are as follows to analyze the results of the basic genetic algorithm:

(1) Generate distribution centers and 50 demand points, 10 proposed charging station nodes, with an initial population of 100;

(2) As the result of Equation 4-1, calculate the fitness function, where;

(3) The basic genetic rule is to select and design a crossover genetic probability of 30% and a mutation probability of 5%;

(4) Store the offspring obtained after mutation or crossover, while also saving the offspring back to the parent population to optimize computational efficiency;

(5) Repeat steps 3 and 4 until the designed termination condition is met. In this article, the designed termination condition is 200 generations;

(6) Compare the offspring with the lowest cost and output the length and order of the proposed charging station and the shortest path, and store them as the optimal solution to the problem.

This article is based on MATLAB R2017a software and uses genetic algorithm to solve the case. Generate an initial population of 100, and after running, the convergence of the algorithm is shown in Figure 1. It can be seen that after approximately 90 iterations, the algorithm begins to converge and tends towards the optimal solution.

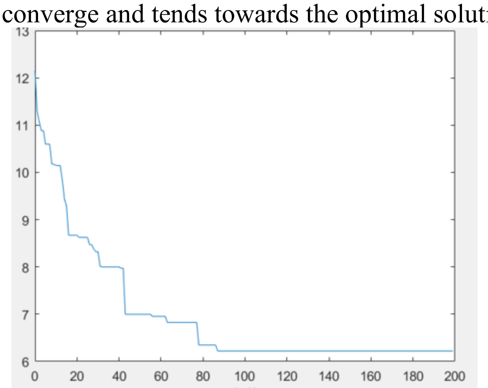


Figure 1 Algorithm Convergence

The optimal solution appears after 163 iterations, and the total driving distance is 1392.6652km.

Finally, the design scheme of combinatorial optimization is shown in Table 4.

Table 4 Final Vehicle Delivery Plan

Plan	Mileage	Load capacity	Plan	Mileage	Load capacity
0-25-27-8-0	39.4082	2520	0-15-2-12-0	118.0956	2520
0-20-24-10-6-0	118.7898	2970	0-1-30-28-0	158.8330	3000
0-3-13-4-0	89.1090	2940	0-16-26-19-0	151.9758	2430
0-21-(C3)-22-9-0	207.5928	2640	0-5-23-(C1)-18-0	203.8643	2790
0-17-29-11-0	78.0555	2640	0-7-(C4)-14-0	226.9412	1920

As can be seen from Table 4, when using genetic algorithm to solve the location path combinatorial optimization problem, a total of 10 new energy vehicles need to be configured, with an average mileage of 139.27 kilometers per vehicle, reaching 87.04%

of the maximum mileage of 160 kilometers for new energy vehicles; The average load capacity is 2637 kilograms, reaching 87.9% of the maximum load capacity of 3 tons, which is a relatively satisfactory state. Compared with the single point delivery plan, the number of vehicles allocated is reduced by 20, the average mileage per vehicle is increased by 92.84 kilometers, and the average load per vehicle is increased by 1758 kilograms. The optimization effect is relatively significant.

5. Conclusion

This article conducts in-depth research on the comprehensive optimization problem of layout planning and optimal path selection of charging station facilities for new energy vehicles in urban distribution systems. This paper summarizes the current situation and research basis of the development of new energy vehicles in urban short and medium distance distribution, and makes a preliminary exploration of the problem of logistics enterprises choosing to establish charging stations by themselves. Finally, with the goal of minimizing the number of charging stations and the shortest transportation mileage, and with logic, travel conditions, time windows, maximum mileage, maximum load capacity and other conditions as constraints, a location path combinatorial optimization model is finally established. At the same time, considering the difficulty of solving multi-objective problems, we have changed the problem of short transportation mileage and fewer charging stations to a single objective optimization problem that comprehensively solves operational costs. On this basis, considering that such practical problems are generally large in scale and involve many factors, a heuristic algorithm is selected to solve the combinatorial optimization model. Taking full advantage of the advantages of the fast convergence speed of genetic algorithm and the ability to jump out of the local optimal search for the global optimal, a genetic algorithm for combinatorial optimization is designed. Finally, an example is used to verify the feasibility and practicability of using genetic algorithm to solve the location routing combinatorial optimization problem of new energy vehicle distribution system.

Due to the fact that this study focuses on the Chinese urban distribution system using new energy vehicles, therefore, the maximum mileage constraint for new energy vehicles is set in the constraint conditions of the model. In other geographical locations or situations where ordinary delivery vehicles coexist with new energy delivery vehicles, further adjustments need to be made to the model based on the cost and operating mileage of different vehicle models.

References

[1] Anagnostopoulou Afroditi et al. Electric Vehicle Routing Problem with Industry Constraints: Trends and Insights for Future Research[J]. Transportation Research Procedia, 2014, 3: 452-459.
[2] Ángel Felipe et al. A heuristic approach for the green vehicle routing problem with multiple technologies and partial recharges[J]. Transportation Research Part E, 2014, 71: 111-128.
[3] Chen et al. Integrated scheduling of zone picking and vehicle routing problem with time windows in the front warehouse mode[J]. Computers & Industrial Engineering, 2022,163: 1-14.
[4] Fu Zhengtang et al. An intelligent green vehicle management system for urban food reliably delivery: A case study of Shanghai, China[J]. Energy, 2022, 257: 1-14.
[5] Dominik Goeke, Michael Schneider. Routing a mixed fleet of electric and conventional vehicles[J]. European Journal of Operational Research, 2015, 245(1): 81-99.

[6] Wu Di. esearch on Vehicle Routing Optimization Based on Genetic Ant Colony Algorithm——Take AN Fresh Agricultural Products Company as an Example[D]. Hebei University of Engineering, 2020.
[7] Weera Chotithammaporn, Ratchasak Sannok, Witthaya Mekhum, Somdech Rungsrisawat, Pitsanu Poonpetpun, Kevin Wongleedee. The Development of Physical Distribution Center in Marketing for Small and Micro Community Enterprise (SMCE) Product in Bangkontee, Samut Songkram[J]. Procedia - Social and Behavioral Sciences,2015,207.
[8] Maximilian Schiffer et al. Vehicle Routing and Location Routing with Intermediate Stops: A Review[J]. Transportation Science, 2019, 53(2): 319-343.
[9] Xu Yuanzhen Comparative Study on Green Vehicle Routing Problem under Joint Delivery Mode [D] Beijing Jiaotong University, 2017.
[10] Dong Yuwen, handsome Green Vehicle Routing Problem Considering Carrying Capacity and Travel Constraints [J] Industrial engineering and Management, 2017,22 (01): 43-49.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230182

Economic Responsibility Audit, Regional Innovation Capacity and Low Carbon Development

Jin SUN[a,1,] Shujia LIN[b], Cuilin WANG[a]
[a]*School of Economics and Management, Lanzhou University of Technology, No. 287, Langongping Road, Qilihe District, Lanzhou City, Gansu Province, China*
[b] *Division of Science and Technology, Beijing Normal University-Hong Kong Baptist University United International College (UIC), Zhuhai, Guangdong, China*

Abstract. Economic responsibility auditing can help promote regional economic development, but whether it helps enhance regional innovation capacity and low carbon development needs to be further tested. This paper examines the objectives of authenticity, legality, and effectiveness of economic responsibility audit. The results show that the increased scrutiny of the authenticity, legality and effectiveness of financial records, policy formulation and implementation, and major project activities of government leaders can definitely enhance the awareness of environmental protection and innovation among cadres, thus improving regional innovation capacity and helping to reduce carbon emissions, achieving the goal of double balance between economic growth and carbon reduction.

Keywords. Economic responsibility audit, regional innovation capacity, low carbon development

1. Introduction

Since the 1950s, countries have pursued efficient production while making extensive use of natural resources, causing environmental problems such as climate warming and biological extinction. As a result, 183 countries signed the Kyoto Agreement in 1998: developed countries assumed carbon reduction obligations from 2005 and developing countries assumed emission reduction obligations from 2012. By 2020, 53 countries have already achieved the carbon peak, and China has officially proposed the goal of "reaching the carbon peak in 2030 and achieving carbon neutrality in 2060"[2], which not only brings certain pressure on China's economic development, but also provides an opportunity to realize the organic combination of "low carbon" and "development",the exploration of new energy technology innovation, low-carbon development of new paths and other development opportunities, in December 2021 for the first time put forward the high-quality development, innovation-driven

---

[1] Corresponding Author, Jin SUN, School of Economics and Management, Lanzhou University of Technology, Lanzhou, Gansu, China. Email: sjjack54331@163.com.
[2] General Debate of the 75th Session of the UN General Assembly in 2020

development strategy and "carbon peaking and carbon neutrality" goals closely linked[3], in August 2022 to achieve the goal of 2030 carbon peak of Low-carbon industrial process reengineering, dual-carbon management decision support system and other initiatives[4]. On the audit side, China emphasized that national audit should better serve in innovative operations and activities[5], and the newly revised Regulations include ecological environmental protection into the key areas of investigation[6] .

The existing studies on audit, innovation activities and low-carbon development show that national audit can promote low-carbon development[1], national audit and exit audit of natural resource assets can help improve regional innovation capacity[2][3], improving R&D intensity and innovation capacity can solve environmental problems[4], but whether strengthening economic responsibility audit of party leaders directly contributes to improving regional innovation capacity and the realization of the "carbon peaking and carbon neutrality" target needs to be specifically tested. Based on the importance and urgency of low-carbon development goals, this paper empirically investigates the mechanism of the three major goals of economic responsibility auditing affecting low-carbon development and the mediating effect of regional innovation capacity using data from 30 provinces in China from 2009-2022. This paper provides useful reference for enriching the research on the role of economic responsibility auditing, developing the application field of innovation capacity, and exploring the realization path of low-carbon development.

2.Theoretical analysis and research hypothesis

At this stage, although China's environmental governance has made great progress and development, to successfully achieve the goal of "carbon peaking and carbon neutrality", we are facing pressure in the practice of green technology innovation and the formulation and implementation of environmental regulation policies, especially in regions with backward economic development, with high environmental pressure and with more prominent contradictions between economic development and environmental protection. Economic responsibility audit not only provide forensic information on fiduciary responsibilities, but also facilitate government to play a better constructive role in maximizing public interests[5].

In low-carbon development, the authenticity and legality objectives of economic responsibility audit can effectively supervise the availability of low-carbon financial funds for allocation, reasonable distribution and accurate records; it can further regulate the formulation and implementation of low-carbon policies and the legal collection and use of low-carbon funds; it can also urge the units with the above problems to rectify and reform. And the effectiveness audit objective of economic responsibility will prompt government leaders to make correct regional economic development goals and rationalize low-carbon development plans, achieve low-carbon economic balance. In the process of helping to achieve the goal of "carbon peaking and carbon neutrality",

---

[3] Guiding Opinions on Promoting High Quality Development of Central Enterprises to Do a Good Job of Carbon Peaking and Carbon Neutral Work

[4] Science and Technology to Support Carbon Neutral Implementation Plan (2022-2030)

[5] Opinions on Auditing to Better Serve the Construction of an Innovative Country and a World Science and Technology Power

[6] Regulations on Economic Responsibility Auditing of Major Leading Party and Government Cadres and Leading Personnel of State-owned Enterprises

economic responsibility audit can promote the management awareness of leaders and the scientific choice of innovation behavior in terms of regional innovation capacity. It can also encourage leaders to exercise their authority, use funds and manage innovation activities in a rational manner, and pay great attention to the economy, efficiency and effectiveness of resource utilization; to enhance innovation capacity and increase economic value added by imitating, absorbing, improving technology and accelerating R&D; to minimize the cost of "trial and error" by means of expert verification[6].

In conclusion, the economic responsibility audit can control the unreal, illegal and ineffective phenomena caused by blind and misplaced investment etc.in innovation behavior in time, stimulate the innovation consciousness of government leaders, form a series of scientific innovation activities, cultivate emerging green industries, and promote the upgrading of low-carbon industries. The hypotheses are proposed:

H1: Economic responsibility audit significantly contribute to low-carbon development.

H2: Economic responsibility audit significantly contribute to regional innovation.

H3: Regional innovation capacity has a mediating role in economic responsibility audit for low-carbon development.

3.Research design and empirical analysis

3.1. Study design

In order to ensure the consistency of the time caliber of the relevant data and the completeness and comprehensiveness of the empirical research data, the study takes the relevant data of 30 Chinese provinces from 2009 to 2020 as the sample. Carbon emissions are obtained from China Carbon Accounting Database, data of economic responsibility audit authenticity and legality indicators are obtained from China Audit Yearbook, innovation capacity index is obtained from China Regional Innovation Capacity Evaluation Report, other data are obtained from CSMAR, data collation and empirical study using spss software and stata software, using regression analysis based on statistical principles of data can determine the correlation between dependent variables and mediating variables, independent variables, so as to establish a regression equation model with good correlation to prove the research hypothesis and predict future changes in the dependent variable.

Carbon emissions(t)measure low-carbon development, the economic responsibility audit authenticity objective is measured by the accounting inaccuracy rate (zs); the legality objective is measured by the violation rate (hf); the factor analysis method is used to construct effectiveness (xy) from tax revenue growth rate, per capita disposable income growth rate, gross product growth rate, budget revenue growth rate, budget expenditure growth rate, and consumption expenditure growth rate. The regional innovation index (cxnl) measures the regional innovation capacity, and the control variables are selected as the number of enterprises (qy), foreign trade (logjckzb) and education and culture(bkzb).Models(1) and (2) verify the impact of economic responsibility audits on low carbon development and regional innovation capacity, and model(3)verifies the mediating effect of regional innovation capacity in the relationship between economic responsibility audits and low carbon development:

$$t=\beta_0+\beta_1 zs+\beta_2 hf+\beta_3 xy+\beta_4 qy+\beta_5 logjckzb+\beta_6 bkzb+\varepsilon \quad (1)$$

$$cxnl=\beta_0+\beta_1 zs+\beta_2 hf+\beta_3 xy+\beta_4 qy+\beta_5 logjckzb+\beta_6 bkzb+\varepsilon \quad (2)$$

$$t=\beta_0+\beta_1 zs+\beta_2 hf+\beta_3 xy+\beta_4 cxnl+\beta_5 qy+\beta_6 logjckzb+\beta_7 bkzb+\varepsilon \quad (3)$$

3.2. Empirical Analysis

Descriptive analysis and correlation analysis are not presented for space reasons. Stepwise regression method is used to test the models from the perspective of authenticity, legitimacy, and effectiveness. The authenticity regression test chart is used as an example, while the legality and effectiveness regression test charts are not shown and textual descriptions are provided.

The multivariate regression of the economic responsibility audit authenticity objective (Table1), $r^2$is 0.039, and the model is significant overall. The test results show that when the rate of accounting inaccuracy (zs) increases, i.e., the audit of authenticity objective (zs) is weak, the carbon emission (t) in model 1 increases significantly at the 5% level with a regression coefficient of 0.025,the hypothesis H1 holds; the regional innovation capacity (cxnl) in model 2 decreases significantly at the 1% level with a coefficient of -0.001, the hypothesis H2 holds. The mediation test of model 3 regional innovation capacity (cxnl) between authenticity target (zs) and carbon emission (t) shows that the significance level of authenticity target (zs) decreases from 5% to 10%, and the regression coefficient decreases from 0.025 to 0.02. The significance level of regional innovation capacity (cxnl) is 10%, the regression coefficient is -4.912, which indicates that the direct effect of economic responsibility audit authenticity objective on low carbon development is reduced, innovation capacity has a partial mediating effect, the H3 holds.

Table 1. Authenticity objective, regional innovation capacity and low carbon development

Variable Type	Variable Name	Carbon emissions (t)	Regional innovation capacity (cxnl)	Carbon emissions (mediating effect)
		Model 1	Model 2	Model 3
Explanatory variables	Authenticity target	0.025**	-0.001***	0.020*
	(zs)	(0.011)	(0.000)	(0.011)
Control variables	Number of enterprises	0.072	0.000	0.072
	(qy)	(0.064)	(0.001)	(0.064)
	Education and Culture	-678.934	-0.672	-682.234
	(bkzb)	(523.529)	(10.665)	(521.706)
	Foreign trade	26.064	-0.937**	21.460
	(logjckzb)	(23.331)	(0.475)	(23.388)
Intermediate variables	Regional innovation			-4.912*
	capacity (cxnl)			(2.709)
	r^2	0.030	0.076	0.039

The multiple regression of the economic responsibility audit legitimacy objective,$r^2$is 0.042, and the model is significant overall. The test results show that when the violation rate (hf) increases.The audit of legality target (hf) is weak, the carbon emission (t) in model 1 increases significantly at the 1% level, with a regression coefficient of 267.847, validating H1; the regional innovation capacity (cxnl) in model 2 decreases significantly at the 1% level, with a coefficient of -11.745, validating H2; model 3, the regional innovation capacity (cxnl) in the mediation test between legality target (hf) and carbon emission (t), the effect of legality target (hf) on carbon emission (t) decreases from 267.847 to 214.255, and the significance level decreases from 1% to 5%, and the regional innovation capacity (cxnl) shows a negative correlation at 10% level with a coefficient of -4.563, verifying H3.

The multiple regression of the economic responsibility audit effectiveness objective ,$r^2$is 0.049 and the model is significant overall. The test results show that: the higher the comprehensive efficiency index (xy), the higher the carbon emissions (t) in model 1 at the 5% level, with a regression coefficient of 135.798, which is at the stage of degradation of environmental quality with increasing income of the environmental Kuznets "inverted U" curve, and is not consistent with the absolute decoupling of carbon emissions with economic growth,H1 is not valid. In model 2, the regional innovation capacity (cxnl) is significantly positively correlated at the 1% level, with a coefficient of 5.397, the good overall efficiency of the local situation, which helps to improve the regional innovation level by putting forward new requirements for industrial upgrading, H2 is valid. In model3, the effect of comprehensive benefit index (xy) on carbon emissions (t) rose from 135.798 to 176.909, the significant degree remains 5%, regional innovation capacity (cxnl) is 1% level negative correlation with a coefficient of -7.617,the comprehensive benefit index (xy) requires high, although carbon emissions (t) appeared to rise, but the mediating effect of cxnl is also obvious,H3 holds.

3.3. Robustness tests

The vif command was used to find that the vif values of the explanatory variables, mediating variables and control variables did not exceed 10, indicating that there was no multicollinearity problem; considering the problem of two-way causal endogeneity, the mediating variables, control variables and explained variables were used to lag one-period data, and winsor2 was used to shrink the tail at the 1% level; considering the problem of endogeneity of omitted variables, the hausman test procedure was used to find that the p-value was 0.0000 to reject the original hypothesis, Fe was correlated at the 1% or 5% level, so the fixed-effects model (fe) was selected to develop the empirical evidence.

The robustness test is conducted by replacing carbon productivity (tc) with carbon emissions (t) , and it passes the test in general with consistent direction and better results. In the main test of authenticity objective (zs), carbon emissions (t), regional innovation capacity (cxnl) and mediating effect are significant at 5%, 1% and 10% levels, while in the robustness test, they are all at 1% level, regional innovation capacity (cxnl) changes from partial to full mediating effect; significance in the main test of legitimacy objective (hf) is significant at 1%, 1%, and 5% levels, while the robustness test is at 1% level, and regional innovation capacity (cxnl) still plays a partial mediating role; the main test of effectiveness objective (xy) is significant at 5%, 1%, and 5% levels, the robustness test is at 1% level, regional innovation capacity (cxnl) changes from partial to full mediating effect. The results of the robustness test prove that the empirical test of this paper is stable.

4.Conclusions and Recommendations

The empirical test results show that strengthening economic responsibility auditing significantly promotes low-carbon development with truthfulness and legality objectives, and legality objective is more effective, but effectiveness objective is in the opposite direction, indicating that China's economic development is positively correlated with carbon emissions and is in the rising stage of the environmental Kuznets "inverted U" curve. The strengthening of economic responsibility audit significantly promotes regional innovation capacity, with legality being the most obvious, followed by effectiveness. Regional innovation capacity has a mediating effect between economic responsibility audit and low carbon development, with the best effectiveness and the second best legality.

Through the analysis, although the authenticity goal to promote low-carbon development is not as obvious as legality, the problem of economic business authenticity has been greatly reduced with China's continuous improvement of economic responsibility auditing, corruption punishment and full audit coverage, the authenticity goal of economic responsibility auditing has been basically achieved. The legitimacy objective is prominent in promoting low-carbon development, mainly because there are still irregularities in economic operation such as misappropriation of special funds. Strengthening the effectiveness goal must help regional innovation capacity and low-carbon economic development.

In order for economic responsibility audit to better contribute to the "carbon peaking and carbon neutrality" goal, we suggest that: on the one hand, auditors continue to pay attention to the economic responsibility audit authenticity goal, and moreover, they should increase the review of legality and effectiveness goals; on the other hand, government leaders should strive to balance the contradiction between economy and environmental protection, and reach the highest point of the environmental Kuznets "inverted U" curve as soon as possible, achieve absolute decoupling of economic growth and carbon emission reduction, enhance regional low-carbon technology research and development capacity to develop low-carbon transition paths on the basis of increased innovation awareness.

References

[1]Li Miao M. Research on national audit and low carbon development [D]. Zhongnan University of Economics and Law, 2019 Sept.

[2]Jinyu Z,Yueying L,Juan L. Do national audits affect regional innovation capacity? --An empirical study based on local government innovation preferences[J]. Journal of Nanjing Audit University,2022 May;19(03):10-19.

[3]Fu Y. Can natural resource exit audit enhance urban innovation capacity [J]. Journal of Guizhou University of Finance and Economics,2022 Jul;(04):52-60.

[4]Sampene Agyemang K,Li C,OtengAgyeman F,Brenya R. Dissipating environmental pollution in the BRICS economies: do urbanization, globalization, energy innovation, and financial development matter?[J].Environmental science and pollution research international,2022 Nov.

[5]Yang Jianrong. Government audit and national economic security in China under economic globalization-an analysis based on emerging classical theory and public fiduciary responsibility[J]. Audit Research,2009(05):9-14.

[6]Guo, General. Theory and practice of latecomer advantage in China[J]. Learning and Exploration,2010(01):132-134.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230183

Research on the Influencing Mechanism of Higher Education Quality in the Context of Big Data*

Tengjun GUO [a, c], Danli LI [b,1], Qingzhong SUN[a,2] and Yulan LIANG [c,3]
[a] Jinan University, Guangzhou, P. R. China. 510632
[b] GuangDong Teachers College of Foreign Languages and Arts, Guangzhou, P. R. China.510640
[c] South China Normal University, Guangzhou, P. R. China. 510631

Abstract. The development of big data has profoundly changed not only the way of human production and life but also the way we learn and educate ourselves. This research conducted an empirical analysis on the influencing mechanism of the quality of higher education using the structural equation model, finding that five factors "Active and Collaborative Learning (ACL)","Student and Faculty Interactive (SFI)","Level of Academic Challenge (LAC)","Supportive Campus Environment (SCE)" and "Enriched Educational Experiences (EEE)" influence the indicators "Knowledge Obtaining (KO)" ,"Ability Obtaining (AO)" and "Value Obtaining (VO)". SCE has the largest impact on KO, while the influence of ACL is relatively weak. On this basis, this paper, by combining the features of big data, presented four methods for improving the quality of higher education: first, enhancing the construction of the education data platform; second, energizing energize education reform via big data; third, accurately identifying the level of academic challenge and selecting the appropriate education model by the aid of the big data technology; and fourth, based on the information technology, building a harmonic teacher-student relationship and peer relationship.

Keywords. Information Technology, Big Data, Higher Education, Education Quality, Learning Engagement

1. Introduction

Big data, featured by convenience and efficiency, has imposed a significant influence on our lives. However, the popularization of big data has also impacted traditional higher education, especially in the aspects of teaching role cognition and education

---

[1] Author Tengjun GUO and Danli LI contributed equally to this work.
[2] Corresponding Authors, Qingzhong SUN, Jinan University, Guangzhou, P. R. China. 510632; E-mail: 1158454922@qq.com and Yulan LIANG, South China Normal University, Guangzhou, P. R. China. 510631; E-mail: 275557821@qq.com.
[3] Author Qingzhong SUN and Yulan LIANG contributed equally to this work.
* This research is supported by the Major Philosophical and Social Science Research Project of the Ministry of Education(21JZD061) and the Guangdong Province Undergraduate Teaching Quality and Teaching Reform Project"An Empirical Study on the Learning Involvement of Hong Kong and Macao Students in Mainland Universities -- A Case study of Jinan University".

methods. In the context of big data, higher education needs corresponding transformation.

Today, research on the quality of higher education worldwide is shifting from the macroscopic level to the microcosmic level, and researchers start to evaluate the effectiveness of education from the aspect of students' individual development. Learning engagement refers to the time and energy students put into effective learning activities and how they perceive the school's support of their study[1][2]. This theory emphasizes student-centered educational philosophy and pays attention to students' real learning experience and their growth and development. It is an important aspect for evaluating students' comprehensive quality and is regarded as a critical theoretical thesis for promoting quality evaluation of higher education to shift from scale orientation to connotation orientation.

This research explores the quality of higher education by studying the relationship between learning engagement and education quality in the context of big data. Besides, by analyzing the factors that influence college students' gains from learning and the acting mechanism of these factors in the context of big data, this paper presents the methods of improving the quality of higher education.

2. Theoretical basis

George officially proposed the theory of learning engagement in 2001, pointing out that learning engagement refers to the time and energy invested by students in academic and effective educational practices, as well as the degree of support provided by the campus environment for students' academic performance. It is the interaction between students' behavioral feelings and the academic environment[3]. Based on the theory of learning Engagement, the United States, Australia, and other countries have carried out surveys of college students' learning situation, in order to provide reference for improving students' learning effectiveness, among which the National Survey of Student Engagement in the United States, NSSE) had the most widespread impact.

On the basis of NSSE, a research team represented by Tsinghua University developed the Survey of Chinese College Students (CCSS). Driven by this project, some experts and scholars have carried out theoretical and practical research on college students' learning engagement, including examining Chinese higher education evaluation from the perspective of learning engagement, promoting CCSS questionnaire indicators and content localization process, and conducting research in combination with specific teaching practices. For sample, Tian Tian et al. built a five-dimensional second-order factor structure model of college students' learning engagement based on CCSS questionnaire data[4]. Sun Dongmei et al. found that interpersonal environment, cultural environment and learning motivation all have significant positive influences on educational harvest[5]. However, previous studies on engagement in learning did not take big data into account as an important factor, which is explored in this study.

From the existing research results on the influence mechanism of learning input on education quality, the research conclusion supports that all dimensions of learning input can predict education quality. For example, Li Xiongying et al. proved that learning input of students in "top-of-the line program" has a significant positive impact on learning harvest[6]. Xing Quanchao pointed out that activities such as "campus environment support degree" and "educational experience richness degree" have a significant impact on undergraduate students' academic performance[7]. Wang Yashun used a multilayer linear model to explore the mechanism of the influence of college students' learning engagement on learning gain. The results show that college students' learning engagement has a higher explanatory rate on learning gain[8]. Liu Yunyun made an empirical analysis and found that hardware, interpersonal and cultural environment were significantly positively correlated with satisfaction at school[9]. Therefore, this paper proposes a research hypothesis: under the background of big data, college students' learning involvement also has a significant positive impact on the quality of higher education.

3. Research design

3.1. Research tool

This research employed CCSS questionnaire as the survey tool. The CCSS questionnaire constructs five-dimensional indicators of learning engagement: Level of Academic Challenge (LAC), Supportive Campus Environment (SCE), Active and Collaborative Learning (ACL), Student and Faculty Interactive (SFI) and Enriched Educational Experiences (EEE). In this paper, these five comprehensive and comparable indicators were selected as the factors for evaluating college students' learning engagement. The quality of education is first reflected in the quality of the education receivers. Therefore, this research focused on students' gains from learning and took the educational outcome-oriented view on value-added education quality evaluation, choosing three indicators, namely Knowledge Obtaining (KO), Value Obtaining (VO) and Ability Obtaining (AO) from indicator "Learning gains from self-report" of the CCSS questionnaire as the indicators to evaluate the education quality.

3.2. Research sample

In this study, samples were collected by grade stratification random sampling method, and questionnaire links were released on wechat and QQ groups in Guangdong universities that widely use big data management, by "Questionnaire Star" tool from March to April 2022. A total of 1764 questionnaires were collected, and 1394 valid questionnaires were obtained after deleting invalid data, with an effective rate of 79%.

4. Empirical analysis

4.1. Testing using the measurement model

4.1.1. Factor analysis

As for the exploratory factors, this research used Spss26.0 software and conducted suitability testing on pretest data with the aid of the KMO value and Bartlett's test of sphericity. The result shows KMO=0.925 and Bartlett's test P=0.00＜0.01, indicating that data in this research can be used for exploratory factor analysis. The method of principal component analysis-direct Oblimin was used to extract factors. After retaining factors with a characteristic value larger than 1 and deleting the items with a communality smaller than 0.04, a loading value smaller than 0.40 and cross-loading of factors, five factors were finally extracted from the learning engagement scale. The final learning engagement scale extracted five factors with a communality between 0.598 and 0.942 and a cumulative variance explained of 62.13%, while the education quality scale extracted three factors with a communality between 0.560 and 0.913 and a cumulative variance explained of 65.91%, both scales reaching a satisfactory level.

As for the confirmatory factors, this research used Mplus7.0 software and conducted confirmatory factor analysis using the maximum likelihood (ML) method, so as to verify the construct validity of the scales. According to the results, the goodness-of-fit indexes of the learning engagement model are $\chi2/df=2.168$, CFI=0.985, TLI=0.959, RMSEA=0.053, and SRMR=0.072; the goodness-of-fit indexes of the education quality model is $\chi2/df=2.586$, CFI=0.912, TLI=0.907, RMSEA=0.049, and SRMR=0.069. This reveals that these two analysis models of the confirmatory factors have a good fitting performance, as shown in Table 1.

Table 1. Goodness-of-Fit Analysis

Goodness-of-Fitting Index	$\chi2/df$	RMSEA	TLI	CFI	SRMR
Learning engagement model	2.168	0.053	0.959	0.985	0.072
Education quality model	2.586	0.049	0.907	0.912	0.069
Fitting criteria	<3	<0.08	>0.90	>0.90	<0.10

4.1.2. Reliability and validity testing

As for reliability, SPSS26.0 was used for reliability testing, and the Cronbach's α coefficient is shown in Table 2. According to the result, the overall Cronbach's α value of the two scales is larger than 0.80, and the Cronbach's α value of each factor is larger than 0.70, indicating good reliability of the questionnaires. At the same time, the combined reliability (CR) of all variables is between 0.726-0.872, larger than 0.60 and reaching a statistically significant level. This reveals a good internal consistency of these two questionnaires and the influence factors, as shown in Table 2.

As for validity, the average variance extracted (AVE) was used to measure the convergent validity. The validity value of each factor is larger than 0.50, which is acceptable. In terms of the discriminant validity, the AVE square root of each factor is overall larger than the correlation coefficient of their respective correlation factors. This indicates a good discriminant validity of the factors and a good correspondence between the measured items, as shown in Table 3.

Table 2. Scale of Reliability and Convergent Validity of the Measurement Model

Scale	Factors	Items	Combined reliability (CR)	α coefficient	Scale α coefficient	Convergent validity
Learning engagement	LAC	9	0.801*	0.796	0.831	0.653
	SCE	10	0.793***	0.783		0.569
	ACL	11	0.760*	0.756		0.461
	SFI	9	0.778**	0.781		0.635
	EEE	12	0.745***	0.822		0.632

Education quality	KO	2	0.872**	0.849		0.547
	AO	10	0.741**	0.786	0.843	0.516
	VO	3	0.726***	0.854		0.620

Note: *** represents p＜0.001, ** represents p＜0.01, and * represents p＜0.05.

Table 3. Discriminant Validity Testing of the Measurement Model

Factors	1. LAC	2. SCE	3. ACL	4. SFI	5. EEE	6. KO	7. AO	8. SCO
1. LAC	0.772							
2. SCE	0.425	0.668						
3. ACL	0.324	0.311	0.751					
4. SFI	0.251	0.365	0.233	0.635				
5. EEE	0.310	0.326	0.224	0.259	0.826			
6. KO	0.266	0.351	0.313	0.367	0.282	0.752		
7. AO	0.339	0.239	0.390	0.392	0.262	0.311	0.846	
8. VO	0.412	0.347	0.408	0.345	0.277	0.372	0.269	0.854

Note: The values on the diagonal are the AVE square root values, and data in the lower triangular matrix are correlated coefficients of the factors.

4.2. Structural equation modeling

The first step of structural equation modeling is to set up the model to be estimated. Based on the research hypothesis, The statistical tool Mplus7.0 was used in this study to build an influence path model for college students' learning gains, as shown in Figure 1. The result of fitting index testing ($\chi 2/df$=2.621, RMSEA=0.045, CFI=0.914, TLI=0.909, SRMR=0.072) reveals that the fitting indexes of the model are in the acceptable scope and the model can well fit data, as shown in Table 4.

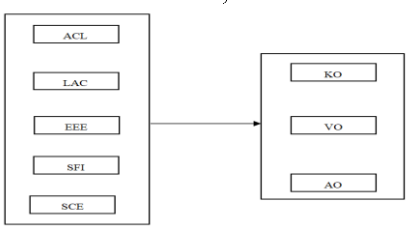


Figure 1. Hypothesis of China's higher education quality influence path model

Table 4. Fitting Effect Evaluation Table of the Structural Model

Fitting index	χ2/df	RMSEA	TLI	CFI	SRMR
Learning engagement model	2.621	0.045	0.909	0.914	0.072

4.3. Hypothesis path testing

The model in Figure 2 clearly shows the factors that influence the quality of higher education, the influence paths, and the structural relationships of these factors. For college students, the five factors of LAC, SCE, ACL, SFI and EEE have a positive influence on KO, VO and AO. According to the results of the data analysis, in this model, except that the paths of LAC and VO, SFI and AO are not significant, there is a statistically significant correlation among 13 paths. There is a significant positive correlation in the paths between LAC and KO (β=0.561, $p<0.05$) and AO (β=0.455, $p<0.001$); there is no significant correlation in the paths between LAC and VO (β=0.496, $p>0.05$). There is a significant positive correlation in the paths between SCE and KO (β=0.652, $p<0.001$), AO (β=0.606, $p<0.05$) and VO (β=0.674, $p<0.01$). The paths between ACL and KO (β=0.295, $p<0.05$), AO (β=0.327, $p<0.001$) and VO (β=0.351, $p<0.05$) are significantly positive. There is a significant positive correlation in the paths between SFI and KO (β=0.373, $p<0.05$) and VO (β=0.432, $p<0.001$), but there is no significant relationship in the paths between SFI and AO (β=0.479, $p>0.05$). The paths between EEE and KO (β=0.599, $p<0.01$), AO (β=0.635, $p<0.001$) and VO (β=0.624, $p<0.001$) have a significant positive correlation. Specific data are shown in Figure 2.

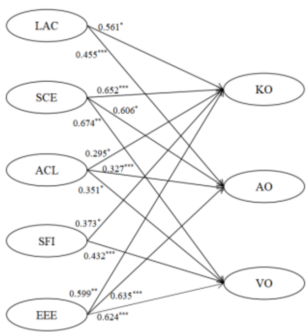


Figure 2. Structural equation model of higher education quality influence factors and influence paths

5. Discussion and suggestions

5.1. Discussion

5.1.1. All the factors of college students' learning engagement have a positive influence on the learning gains

According to the results of data analysis, college students' learning engagement has a significantly positive influence on their learning gains. In other words, students taking more challenging courses, spending more time and energy in their learning, increasing high-quality interactions with their peers and teachers, and actively taking part in various educational practices, all these activities can effectively improve their abilities such as synthetic analysis and critical thinking, and help the college students to have a better understanding of themselves and the world.

5.1.2. The influence of the factors of college students' learning engagement on their learning gains is different

The empirical research suggests that SCE is the most important factor that influences the quality of higher education, having an effect value above 0.6 with KO, VO and AO.

The second most influential factor is EEE, which has an effect value above 0.6 with AO and VO and an effect value of 0.59 with KO. In comparison, ACL and SFI have a relatively weak influence on college students' learning gains, with an influence value below 0.5.

5.1.3. SCE has the largest influence on college students' learning gains, while ACL has the smallest influence

The empirical research indicates that the influence path of SCE is the most significant factor for the quality of higher education. The main reason is that the campus environment has a powerful educational function and can unconsciously influence students' choice of values, way of thinking and behavioral tendencies[10]. As to ACL, analysis reveals a small influence coefficient between ACL and KO, VO and AO. This could be attributed to the lecture-based approach typically used by college instructors, which offers limited chances for students to engage in oral presentations, ask questions, or collaborate with peers. Consequently, ACL may not effectively enhance students' learning gains[11].

5.2. Suggestions

Today, the information technology represented by big data is developing quickly, which not only profoundly influences the reform and development of all sectors of society but also provides unprecedented opportunities for innovating the methods of higher education. Based on the results of the empirical study discussed above, we have the following suggestions of improving the quality of higher education by combining the features and advantages of information technology in the era of big data.

5.2.1. To enhance the construction of the education data platform

First, colleges need to set up an education data platform that integrates massive educational resources and information. In building this data platform, colleges should collect a magnitude of educational data and resources by screening and integrating data information and broadening the channels of data sources. At the same time, it is suggested colleges strengthen cooperation with enterprises, with the help of modern information technology, to set up a data platform integrating educational administration, student management, teaching and academic research. Furthermore, by using cloud computing technology, colleges can classify and analyze data on the platform, thereby gaining insight into the features of different student groups and seeking a basis for the innovation of education methods.

5.2.2. To energize education reform via big data

In educational practices, colleges can get an understanding of the number of participants and students' feedback on the practices with the aid of the big data platform, learn the advanced experience that helps enrich their educational pattern and content and keep broadening their thinking in the learning process. In social practices, colleges can also use big data technology to record the whole process of the out-of-school education practice in real time through various we-media platforms. They can also adopt a unified educational information-sharing system to accurately reflect students' participation in the practices and their feedback, in an attempt to continuously improve the educational effect of the practices.

5.2.3. To accurately identify the level of academic challenge and select the appropriate education model depending on the big data technology

College administrators and teachers can collect and manage information such as the click rate of teaching content, the keyword search rate and the applicable level corresponding to students' abilities by using the big data platform. This helps the colleges to select appropriate educational content and reduce misguide from disqualified and wrong educational resources quickly and accurately. Colleges can also use the big data platform to search educational patterns of other colleges with the same major features and the same core disciplines, to learn from their beneficial experience and develop their respective educational patterns by combining their situations.

5.2.4. To build a harmonic teacher-student relationship and peer relationship based on information technology

In terms of the teacher-student relationship, teachers can use the big data platform to observe and understand each student's real situation of study and life and their mental state. Through data comparison and analysis, teachers can study the type of changes in students' thoughts, thereby communicating with students in a more targeted manner. As to the peer relationship, colleges can take full use of information technology to break boundaries in time and space, encourage more communication between students and create a harmonic and friendly peer atmosphere. This will further improve the effectiveness of collaborative learning and feasibly raise education quality.

6. Conclusions and limitations

Based on a sample of 1394 college students in China, this study proposed and tested the influence of students' learning involvement on the quality of college education in the era of big data. The results show that each dimension of college students' learning engagement has a positive influence on learning harvest, and the influence is different. Campus environment support has the biggest influence on college students' learning harvest, while active cooperative learning has the least influence. In conclusion, this study confirms that in the era of big data, students' learning involvement has a significant positive impact on the quality of higher education.

This study has some limitations that could be addressed in future studies. First, the sample size of this study is limited. More participants with diversified backgrounds could be included in future studies. Secondly, the data in the current study were collected with self-report and cross section. Although the measurements were widely used in previous studies and were proved to be reliable and valid in the current study, might still have led to a subjective bias and difficulties to track the study. Consequently, it is recommended for future studies to adopt multiple evaluation methods and data (e.g., field visits, tracking studies) with reliable measurements to evaluate the learning engagement of college students and the quality of higher education.

References

[1] KUH G D, “The National Survey of Student Engagement: Conceptual and Empirical Foundations”, New Directions for Institutional Research, vol.141,pp. 5-20,2009.

[2] ASTIN A W, ANTONIO A L. "Assessment for excellence : the philosophy and practice of assessment and evaluation in higher education”. Washington: Rowman & Littlefield Publishers pp. 21-22,2012.

[3][6] Li Xiongying, Qin Xiaoqing, “A Study on the Relationship between Learning Engagement and Learning Gain of Students in ‘Top-notch Program’ -- Also on the Promotion of Deep Learning for College Students”, Jiangsu Higher Education, vol.12, pp. 102-108, 2019.

[4] Tian Tian, Zhang Linghua, “Reresearch on the Learning Engagement Structure of College Students based on CCSS2016 Questionnaire”,Higher Education Research in Heilongjiang Province, vol.06, pp 119-125, 2018.

[5] Sun Dongmei, Liu Yunyun, “Research on the Influence Mechanism of Education Quality in research Universities: Based on the analysis of CCSS survey data of L University”, Fudan Education Forum, vol.06, pp. 56-64, 2018..

[7] Xing Quanchao, “An empirical study on the relationship between academic engagement and academic performance of undergraduate students based on CCSS survey ”, Chongqing University, pp.25, 2015.

[8] Wang Ya-shuang, “An empirical study on College students' Learning Engagement -- Based on the data of the National College Students' Learning Survey in 2012”, Higher education research in China,vol.01, pp. 32-36, 2013.

[9] Liu Yunyun, “Research on Influencing factors of college satisfaction based on CCSS survey data of L University”, Lanzhou University, pp.42, 2018.

[10] Wang Peipei, Yu Anlong, “Construction of Learning Evaluation System of Ideological and Political Courses in Colleges and Universities -- Exploration of Value-oriented innovation Path Based on moral cultivation”, Journal of Tianjin Normal University (Social Science Edition),vol.01, pp. 8-13, 2023.

[11] Li Yangyang, Liu Sannu-ya, Chen Zeng-zhao, FANG Jing, HE Xiu-ling, “Research on Collaborative Learning Model Construction from the perspective of Learning Space Integration”, Distance Education in China, vol.03, pp.42-50+60, 2022.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230184

Multi-Criteria Decision-Making for Investment Portfolio Selection in Thailand's Stock Market

Pornpimol Chaiwuttisak[a,1]
[a] *Department of Statistics, School of Science, King Mongkut's Institute of Technology Ladkrabang, Bangkok, 10520 THAILAND*

Abstract. This research aimed to determine the relative importance of four criteria for decision-making by using each one with a popular decision-making method and evaluating the outcomes and using these criteria to provide ranked alternatives (according to each criterion's relative importance) for making an investment decision. The four criteria were obtained from a comprehensive literature review related to securities investment. The investment data analyzed were past investment data on trading securities under the Energy and Utilities category of the SET50 index in the Stock Exchange of Thailand. The analysis was done through an Analytic Hierarchy Process (AHP) and a Technique for Order of Preference by Similarity to the Ideal Solution (TOPSIS). Opinions of three experts with experience in giving securities investment advice were collected and arranged into pair-wise comparison matrices that were used in AHP. AHP and TOPSIS calculations were done in Microsoft Excel. The results of the study show that the most important criterion was financial fundamentals with a weight of 44.59%; the second rank criterion was technical factors with a weight of 20.15%; the third-rank criterion was risk factors with a weight of 19.64%; and the last rank criterion was fundamentals of structure and sustainable development with a weight of 15.62%. In addition, the outcome of security ranking by TOPSIS and the past security ranking data were significantly similar as analyzed by a hypothesis statistical test with two dependent samples.

Keywords. Analytic Hierarchy Process, Multi-criteria Decision Making, Stock Investment, Technique for Order Preference by Similarity to Ideal Solution

1. Introduction

Financial investment is the dedication of an investor's money to buy various securities in order to generate income in the future. The revenue or return depends on the conditions and agreements of the securities. Generally, under a low-interest rate state, assets will be able to generate a higher revenue than the interest that must be paid. Thus, shareholders can make a lot of profit using low personal investment. On the other hand, if investing in high-interest rates, the interest burden will rise and will make less profit to shareholders. Securities have different risks. There are many forms of returns from investment in securities, such as capital gain, dividend, and rights offering.

[1] Corresponding Author, Pornpimol Chaiwuttisak, King Mongkut's Institute of Technology Ladkrabang, Bangkok, 10520 THAILAND; E-mail: pornpimol.ch@kmitl.ac.th.

The fluctuations in stock prices may be the result of several factors. Investors in securities need to have basic knowledge, such as reading financial statements, understanding the basic factors of the company, and studying techniques for stock analysis. Therefore, investors need to study various information before deciding to invest and to find new methods that can make a profit as needed. However, most investors make investment decisions based mainly on their intuition [1], which requires a reasonable decision. A good decision depends on the decision-making processes. Important decisions require a process with steps relying on the reason to lead to the correct and successful path. The important step in the decision-making process is to determine the importance of the criteria that will affect the choice of options to invest, especially for new investors or small investors who still lack experience [2].

[3] applied non-parametric statistical analysis to determine ranking criteria, ANP Multi-Fuzzy technique to determine the weights. The Fuzzy TOPSIS technique was then used to rank the top 50 stocks. Some experts believe that 80 percent of ranking results are based on the first three factors that carry more weight. In addition, the results of expert surveys about being ranked in the rankings show that the results are very close to reality. [17] have proposed a model for evaluating India's best pharmaceutical stocks. using Multi-criteria Decision Making (MCDM). The weights are calculated from pairwise comparisons based on the best and worst criteria. The main criteria are the likelihood of bankruptcy score, Z-score, total debit bonus, cash per share, book value per share, operating profit margin, price-earnings ratio, and revenue. As a result, Wipro is a stable stock, while Tech Mahindra became the stock with the highest price increase.

The objectives of the study are to study criteria used in making investment decisions in energy and utility sectors in the Stock Exchange of Thailand, to determine the weight of the criteria and prioritize the criteria by using the Analytic Hierarchy Process (AHP) method, and to apply AHP and Technique for Order of Preference by Similarity to the Ideal Solution (TOPSIS) in making investment decisions in energy and utility sectors.

1.1. Analytic Hierarchy Process (AHP)

The Analytic Hierarchy Process is the widely popular method of multicriteria decision-making techniques. It is the process to support a complex decision, which can be broken down into a set of criteria relating to the overall goal by the pairwise comparison method and calculating a rating of criteria and alternatives [4, 5]. Moreover, it can help to make good decisions in situations of several choices [6, 7, 8].

There are three parts: the goal, all of the possible solutions, called alternatives, and the criteria on which decision-makers will judge the alternatives. The step-by-step approach for AHP is followed as:

Step 1: Define the problem and criteria.

Step 2: Define alternatives.

Step 3: Establish priority amongst criteria and alternatives using pairwise comparisons.

Step 4: Check consistency amongst the pairwise comparisons.

To calculate consistency for scoring criteria by pairwise comparison. It takes the sum of the values for each criterion in the vertical row. Each row is multiplied by the sum of the mean values in each horizontal row, then add the multiplied results obtained together. as in equation (1).

$$AW = \lambda_{\max} W \tag{1}$$

Where A denotes a reciprocal matrix by a numerical value normalized to 1.

W denotes eigenvectors that show the relative importance of criteria of the same hierarchy.

$\lambda_{\max}$ denotes maximum eigenvalue.

In the next step, to validate the results of the AHP, the Consistency Ratio (CR) is calculated using the formula, CR = CI/RI in which the Consistency Index (CI) is, in turn, measured through the equation (2):

$$C.I. = \frac{(\lambda_{\max} - \mathrm{n})}{(\mathrm{n}-1)} \tag{2}$$

Random Consistency Index (R.I) is obtained by experimentally sampling from 64,000 sample matrices proposed by Saaty [4] as shown in Table 1.

Table 1. The value of the Random Consistency Index

Dimension	1	2	3	4	5	6	7	8	9	10
RI	0.00	0.00	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49

Step 5: Evaluate relative weights from the pairwise comparisons and get the calculated overall priorities for the alternatives.

1.2. Technique for Order Preference by Similarity to Ideal Solution (TOPSIS)

The Technique for Order Preference by Similarity to Ideal Solution is one of the multicriteria decision-making techniques. [9] stated that the main concept is based on an ideal solution technique. The alternative which is close to the positive ideal solution and far from the negative ideal solution is the best. The importance weights of the criterion are defined by the expert's opinion numerically and the ranking results can be obtained corresponding to the importance weights of the criteria.

Step 1: Construct the decision matrix and determine the weights of the criteria.

Let $X = \left(x_{ij}\right)$ be a decision matrix and $W = \left[w_1, w_2, \ldots, w_n\right]$ be a weight vector, where $x_{ij} \in \Re, w_j \in \Re$

Step 2: Calculate the normalized decision matrix.

The normalized decision matrix is defined in (3).

$$r_{ij} = x_{ij} \Big/ \sqrt{\sum x_{ij}^2} \tag{3}$$

Step 3: Calculate the weighted normalized decision matrix

The weighted normalized decision matrix is defined in (4).

$$v_{ij} = w_j r_{ij} \tag{4}$$

Step 4: Determine the positive idea and negative ideal solutions.

Positive ideal solution A^+ has the form shown in (5):

$$A^+ = \{v_1^+, v_2^+, v_3^+, \dots v_j^+, \dots v_n^+\} \tag{5}$$

$$= \{(\max_i v_{ij} \,|\, j \in j_1), (\min_i v_{ij} \,|\, j \in j_2), i = 1, 2, 3, \dots, m\}$$

Negative ideal solution A^- has the form shown in (6):

$$A^- = \{v_1^-, v_2^-, v_3^-, \dots v_j^-, \dots v_n^-\} \tag{6}$$

$$= \{(\min_i v_{ij} \,|\, j \in j_1), (\max_i v_{ij} \,|\, j \in j_2), i = 1, 2, 3, \dots, m\}$$

Step 5: Calculate the separation measure from the positive ideal solution and negative ideal solution.

The separation of each alternative from the positive ideal solution is given in (7).

$$s^+ = \sqrt{\sum_{j=1}^{n} \left(v_{ij} - Aj^+\right)^2}, i = 1, 2, 3, \dots, m \tag{7}$$

The separation of each alternative from the negative ideal solution is given in (8).

$$s^- = \sqrt{\sum_{j=1}^{n} \left(v_{ij} - Aj^-\right)^2}, i = 1, 2, 3, \dots, m \tag{8}$$

Step 6: Calculate the relative closeness to the positive ideal solution.

The relative closeness to the positive ideal solution is defined in (9).

$$c_i^+ = \frac{s_i^-}{\left(s_i^+ + s_i^-\right)}, i = 1, 2, 3, \dots, m \tag{9}$$

Step 7: Rank the preference order or select the alternative closest to 1.

2. Research Methodology

The data used in this study are divided into 3 parts: 1) primary data collected from experts 2) secondary data collected from industrial sector securities, energy, and utilities listed on the Stock Exchange of Thailand. A total of 12 securities are shown in Table 2 and 3) technical analysis data from the efin Stock Pick Up® program. Data is collected from 2021 until 3 March 2022.

Table 2. The name of the securities company in the research

No	Company Name	Abbreviation
1	Banpu Public Company Limited	BANPU
2	B.Grimm Power Public Company Limited	BGRIM
3	Banpu Power Public Company Limited	BPP
4	Energy Absolute Public Company Limited	EA
5	Electricity Generating Public Company Limited	EGC
6	Global Power Synergy Public Company Limited	GPSC
7	Gulf Energy Development Public Company Limited	GULF
8	IRPC Public Company Limited	IRPC
9	PTT Public Company Limited	PTT
10	PTT Exploration and Production Public Company Limited	PTTEP
11	RATCH Group Public Company Limited	RATCH
12	Thai Oil Public Company Limited	TOP

The main criteria have been selected to use in a decision in securities investment consisting of four main criteria and 11 sub-criteria as shown in Table 3. The experts, who have experience in advising securities investment, have been selected to evaluate the relative importance scores. There are three experts, consisting of the first expert working as a Wealth Relationship Manager, the second expert as a Senior Director of Human Resources, and the last expert working as Dealer & Products (Global Markets). Those experts have the following qualifications: 1) Able to give advice for trading or investing in the stock market. 2) Have basic knowledge of finance and investment in several types of financial securities, including the ethic of conduct for giving appropriate investment advice.

Table 3. Main criteria and sub-criteria used in making investment decisions in securities.

Main Criteria	Sub-criteria	Related Research	Description
Financial Infrastructure	Return on Equity Price to Earnings Ratio Price to Book Value Ratio	[10,11,12,13]	All sources of funds provided by the business for its operations, both short-term and long-term sources of funds.
Technical Indicators	Oscillator Relative Strength Index Simple Moving Average	[1,14]	Tools Selection helps make a profit from the market which can help to spot signals. or an opportunity to make a buy signal or a sell signal
Basic Structure and Sustainable Development	Governance Corporate Image	[15]	Concrete practice for sustainable growth goals of the business
Investment Risk Factors	Beta Standard Deviation Liquidity Ratio	[9,16,17]	Changes that cause the actual yield to deviate from the expected rate of return.

3. Result

All criteria and alternatives mentioned above have been used to formulate a hierarchical diagram of decision-making. The objective at level 1 is to select securities in the energy and utilities group. The main criteria are at level 2, sub-criteria are at level 3, and alternatives are at level 4. Then, the geometric mean of the pairwise comparisons of criteria obtained by the three experts was calculated as in Table 4, dividing each pairwise comparison by the sum of the numbers in the vertical column of the main criteria. It is normalized by having the sum of the main criteria in each column equal to 1.00 and calculating the importance of the main criteria by calculating the mean of the main criteria in each row shown in Table 5. It can be seen that the top priority criterion in making a decision to invest in securities is the financial infrastructure, followed by technical indicator, investment risk factor, and basic structural & sustainable development respectively.

Table 4. Consistency analysis of the main criteria

Main Criteria	Financial Infrastructure	Basic structure and sustainable development	Technical Indicators	Investment Risk Factors	Eigenvector	Product	Product/ Eigenvectors
Financial Infrastructure	1.0000	2.7589	3.3019	1.6134	0.4459	1.8591	4.1694
Basic Structure and Sustainable Development	0.3624	1.0000	0.7306	0.8707	0.1562	0.6360	4.0727
Technical Indicators	0.3018	1.3821	1.0000	1.3572	0.2015	0.8185	4.0620
Investment Risk Factors	0.6136	1.1447	0.7368	1.0000	0.1964	0.7973	4.0586
			λ_{max}				4.0907
	C.I.		0.0302		C.R.		0.0336

Table 5. Weights of the main criteria.

Main Criteria	Weights	Priority Orders
1 Financial Infrastructure	0.4459	1
2 Basic Structure and Sustainable Development	0.1562	4
3 Technical Indicators	0.2015	2
4 Investment Risk Factors	0.1964	3

After that, the maximum Eigenvector (λ_{max}) was calculated by multiplying the product of eigenvalues from Table 5 in each row by the mean of pairwise comparisons by experts of all criteria from Table 4, then calculating the Consistency Index (C.I.) by subtracting the number of criteria, which in this research is 4, dividing by the number of criteria minus 1, then calculating the Consistency Ratio (C.R.) from C.I. divided by the Random Consistency Index (R.I.), which is a sampling of 64,000 samples from the matrix. In this study, there were 4 main criteria for sampling. Therefore, the value is 0.90, indicating that the Consistency Ratio: C.R.) has a value of 0.0336, which is not more than the specified value, which is 0.10, as shown in Table 4.

In the same way, the researcher used the results of comparing the sub-criteria significance from all experts to calculate the average significance weight. as well as checking the consistency ratio, which must not exceed 0.10. The calculation results of the weighting of the sub-criteria under the financial infrastructure criteria show that the first-ordered priority in deciding to invest in securities is the Price to Earnings Ratio representing 26.27 percent. The second-ordered priority is the Price to Book Value Ratio. representing 10.18 percent. The third-ordered priority is the return on equity ratio. accounted for 8.13 percent.

The result of the significant weight of sub-criteria under the Basic Structure and Sustainability Criteria shows that top priority in deciding to invest in securities is the image of the company, representing 9.22 percent, and good governance representing 6.40 percent. The results of calculating the weight of the sub-criteria under the technical indicators show that the top priority is the simple moving average (9.33 percent), followed by the Stochastic Oscillator (5.91 percent), and a relative strength index (4.91 percent). The results of calculating the weight of the sub-criteria under the risk factor, it was found that the liquidity ratio is 9.19 percent, the beta value is 5.51 percent, and the standard deviation accounted for 4.94 percent.

Then, TOPSIS was applied to help rank securities. All 11 sub-criteria were collected as follows: Return on Equity Ratio, Price to Earnings Ratio, Market Price to Book Value Ratio, Good Governance, Company Image, Relative Strength Index, Stochastic Oscillator, Simple Moving. Average, Beta, Standard Deviation, and Liquidity Ratio. They are determined as the benefit criteria (+) and cost criteria (-) and determining the final weight derived from the hierarchical analysis process (AHP) as in Table 6.

Table 6. AHP weights and benefit/cost of all 11 sub-criteria

Main Criteria	Sub Criteria	Symbol	AHP Weights	Relative Weights	Benefit/Cost Criteria
Financial Infrastructure	Return Ratio to Shareholders	C1	0.0813	0.1824	+
	Price Ratio Per Profit	C2	0.2627	0.5893	+
	Accounting Market Price Ratio	C3	0.1018	0.2283	-
Basic structure and sustainable development	Governance	C4	0.0640	0.4095	+
	Image	C5	0.0922	0.5905	+
Technical Indicators	Relative Strength Index	C6	0.0491	0.2436	+
	Stochastic Oscillator	C7	0.0591	0.2932	-
	Simple Moving Average	C8	0.0933	0.4632	+
Investment Risk Factors	Beta	C9	0.0551	0.2804	-
	Standard Deviation	C10	0.0494	0.2515	-
	Liquidity Ratio	C11	0.0919	0.4681	+

The researcher collected data for all 11 sub-criteria as mentioned above, of 12 securities in the energy and utilities category of the SET50 listed on the Stock Exchange of Thailand in 2021, between the first quarter of 2021 and the fourth quarter of 2021, all 12 securities are alternatives for making investment decisions. and then create a 12 x 11 decision matrix to calculate the stock rankings. Then adjust the data by vector normalization method and calculate the standard weight by multiplying the standard of each alternative and each criterion with the AHP weight.

After that, determining the ideal solution matrix of the positive ideal solution with the highest value of each criterion (V^+) and the negative ideal solution with the lowest value of each criterion (V^-) were calculated. Then calculate an alternative distance from the positive ideal solution (S_i^+) and an alternative distance from the negative ideal solution (S_i^-) and find the consistency value to be a positive ideal solution weight (C_i^+). Table 7 shows the values S_i^+, S_i^- and C_i^+ of securities used in making investment decisions in the first quarter of 2021.

Table 7. S_i^+, S_i^- and C_i^+ of securities used in making investment decisions in the first quarter of 2021

Securities	S_i^+	S_i^-	$S_i^+ + S_i^-$	C_i^+	Order
BANPU	0.1709	0.0860	0.26	0.3346	11
BGRIM	0.1065	0.1169	0.22	0.5233	2
BPP	0.1560	0.0862	0.24	0.3558	8
EA	0.1356	0.0965	0.23	0.4157	3
EGCO	0.1725	0.1048	0.28	0.3781	6
GPSC	0.1422	0.0897	0.23	0.3867	4
GULF	0.0802	0.1692	0.25	0.6786	1
IRPC	0.1679	0.0844	0.25	0.3346	12
PTT	0.1662	0.0852	0.25	0.3388	10
PTTEP	0.1562	0.0857	0.24	0.3543	9
RATCH	0.1564	0.0882	0.24	0.3606	7
TOP	0.1581	0.0981	0.26	0.3830	5

The Sign Test for Two Related Samples is used to compare actual rankings of securities based on the percentage change of securities and securities rankings using TOPSIS.

Statistical Hypothesis

H_0: The securities ranked by TOPSIS did not differ from the actual ranks.

H_1: The securities ranked by TOPSIS differ from the actual ranks.

The results of the hypothesis test obtained from the IBM SPSS analysis between the securities ranked by TOPSIS and the actual ranks of securities, at a significance level of 0.05. The first quarter of 2021 has a p-value of 0.388, which is higher than the significance level. Therefore, the null hypothesis cannot be rejected. Namely, the securities ranked by using the TOPSIS method and the actual ranks are not different. As same as, the securities ranked by using the TOPSIS method and the actual ranks in the other quarters are not different.

4. Conclusions

The ranking of the main criteria for stock market investment on the AHP. It was obtained by experts. The experts have given importance to the main criteria in descending order as follows: First, the financial infrastructure (0.4459), followed by the technical indicators (0.2015), the investment risk factor (0.1964), and basic structure and sustainable development (0.1562).

The researcher used the TOPSIS method to analyze alternative priorities of 12 securities in the Stock Exchange of Thailand during the 1st - 4th quarter of the year 2021, using statistics to test using the Sign Test for Two Related Samples to test for consistency. between the securities ranks using the TOPSIS method and the actual ranks of securities at the significance level of 0.05. The results showed that there was no difference between the two groups, ranks by TOPSIS are no different from actual ranks of securities.

References

[1] Soniwan RG, Ratnawati A, Irawan T. Analysis of Decision Making Investment Stock Using Analysis Hierarchy Process (AHP). International Journal of Business and Management Review. 2017; 5(10): 31-39.

[2] Opasanon S. Multiple Criteria Decision Making. Journal of Business Administration. 2003; 35(139): 4 – 7.

[3] Kimiagari AM, Saleh ES. Ranking Tehran's Stock Exchange Top Fifty Stocks Using Fundamental Indexes and Fuzzy TOPSIS. Engineering, Technology & Applied Science Research. 2017; 7(4): 1863 – 1869.

[4] Saaty TL. The Analytic Hierarchy Process. New York: McGraw-Hill; 1980.

[5] Saaty TL. Multicriteria Decision Making: The Analytic Hierarchy Process. Pittsburgh: RWS Publications; 1991.

[6] Ghodsypour SH, O'Brien C. A Decision Support System for Supplier Selection using an Integrated Analytic Hierarchy Process and Linear Programming. International Journal of Production Economics. 1998; 56-57: 199 - 212.

[7] Benyoucef L, Ding H, Xie X. Supplier Selection Problem: Selection Criteria and Methods. 2013. Research Report.

[8] Ho W, Bennett D, Mak KL, Chuah KB, Lee CKM, Hall M J. Strategic Logistics Outsourcing: An Integrated QFD and AHP Approach. Proceedings of the 2009 IEEE International Conference on Industrial Engineering Management. 2009; 1434 – 1438 p.

[9] Ilhan Z, Tatlıdil H. Multiple criteria decision analysis methodologies and an application to the Istanbul Stock Exchange. International Journal of Operational Research. 2016; 26(2): 180-195.

[10] Madi EN., Tap AO Md. Fuzzy TOPSIS method in the selection of investment boards by incorporating operational risks. Proceedings of the World Congress on Engineering; 2011, July 6 – 8; London, U.K.

[11] Raei R, Jahromi MB. Portfolio optimization using a hybrid of fuzzy ANP, VIKOR and TOPSIS. Management Science Letters. 2012; 2(7): 2473-2484.

[12] Ece O, Uludag A. Applicability of Fuzzy TOPSIS Method in Optimal Portfolio Selection and an Application in BIST. International journal of economics and finance. 2017; 9: 107-127.

[13] Saleh ES, Kimiagari AM. Ranking Tehran's Stock Exchange Top Fifty Stocks Using Fundamental Indexes and Fuzzy TOPSIS. Engineering. Technology & Applied Science Research. 2017; 7: 1863-1869.

[14] Pumalee P, Teeravaraprug J. The Use of Analytic Network Process to Analyze Significant Factors of Short-Term Investment of Retail Investors in Stock Exchange of Thailand. MUT Journal of Business Administration. 2013; 1(10): 80 – 91.

[15] Abbassi M, Ashrafi M, Tashnizi ES. Selecting balanced portfolios of R&D projects with interdependencies: A cross-entropy based methodology. Technovation. 2014; 34(1): 54 – 63.

[16] Janani MH, Ehsanifar M, Bakhtiarnezhad S. Selection of Portfolio by using Multi Attributed Decision Making (Tehran Stock Exchange). American Journal of Scientific Research. 2012; 44(2): 15 – 29.

[17] Krishna CS, Niyas K, Kumar PAR. Model for Choosing Best Stock among A Set of Stock in Stock Market. International Journal of Advances in Science Engineering and Technology. 2018; 6(1): 39 – 47.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230185

An Empirical Study on the Development Routes of Library Under the Cultural Digitization Strategy

Yan Changmao[a], Rong Peng[b1]
[a]*Xiangnan University Library, Chenzhou-423000,China*
[b]*Bocai Experimental Middle School Attached to Hunan Normal University, China*

Abstract. With the rapid development of information technology, there is an emergence of the advancement in cultural digitalization and digital industrialization. By analyzing the implications of the value and technical routes of a cultural digitization strategy, the present study proposed a cultural digitization strategy that transforms the related works, knowledge and wisdom of cultural producers into quantifiable, traceable and tradable digital assets. Additionally, it promotes a more profound level of processing, including the processing and service of cultural resources, which further extends the supply chain of cultural digital resources, while promoting the integration and development within business sectors. This can assist the upgradation and transformation of the cultural resources to ensure an improvement of the digital service levels of cultural resources, along with stimulating new vitality norms of the cultural resources. Further, it can be a motivation factor to promote the creative development of a national culture. Libraries are cultural hub enriched with literature within diverse audience to drive the information circulation. The implementation of a cultural digitization strategy can effectively promote the transformation and upgradation of the library elements by providing a solid channel to focus on libraries as an institution resulting in strategic positions of cultural power in a changing global environment. By focusing on the key tasks and long term goals of cultural digitization, the active adoption and implementation of "enabling digital technology applications by integration and innovation" can be achieved.

Keywords. Cultural digitization; Library; Implementation lines; Development routes

1. Introduction

Over the last 5 years, with a rapid development of 5G/6G, artificial intelligence, AR/VA/MR, block-chain and other information technologies, and the integrated development of science and culture, there is a gradual emergence of cultural

---

[1] Corresponding author, Peng Rong, Bocai Experimental Middle School Attached to Hunan Normal University, China; Email: 794008994@qq.com
It is the research result of The Key Cultivation Base for "The 14th Five-Year Plan" of Educational and Scientific Research (Lifelong Education Research Base(Fundamental Theory Area)) in Hunan Province (XJK22ZDJD58). The present study was supported by the Special Project of Intellectual Property Strategy Promotion in Hunan Province, 2021 (Research on the Evaluation Index System of Intellectual Property Promoting High-quality Development in Hunan and its Countermeasures) (Project No. : 2021F006K).

digitalization and digital industrialization. With the constant unlocking of the integration of cultural resources, there is a constant evolution of new forms of cultural industry and new scenes of cultural consumption. With an acceleration of the developments in the Chinese digital culture industry, cultural digitization has become a national strategy [1]. The digitization of culture is an important means for the "creative transformation and innovative development" of culture, and it is considered an appropriate route to build a strong socialist culture. Additionally, it enables the support in the development of a digital economy, thereby building a cultural power and supporting the rejuvenation of China as a nation. As a cultural place for the public to consult literature and obtain information and spread knowledge, there is a need to develop high-quality libraries that can adapt to the general trend of cultural digital development, and strengthen the development of new cultural formats. Currently, there is a need to transform and improve the traditional cultural formats, enhance the supply of high-quality digital cultural products, and create sustainable digital cultural samples by the libraries. These issues are some of the practical subjects that need to be contemplated by the librarians.

2. Connotation, value and technological route of cultural digitization strategy

2.1 Evolution process of a national strategy ascending with cultural digitization

Tracing its origin, China's cultural digitization construction work started since past ten years. In 2012, the Outline of Cultural Reform and Development Plan during the 12th Five-Year plan proposed the implementation of the construction project of "cultural digitization"[2], pointing out that the comprehensive digitization of cultural resources, cultural production and cultural communication should be promoted to boost the integration of culture and technology. The need to promote the digitization of cultural resources has also been drafted into the report of the 18th National Congress of the CPC [3] in the same year. Since then, it has opened a new stage, wherein the culture and technology continue to accelerate the integration and development process. In 2019, the Chinese Ministry of Science and Technology and other six departments jointly issued the "Guiding Opinions on Promoting the Deep Integration of Culture and Technology", that presented clear guiding opinions on the deep integration of culture and science and technology in key fields such as industrialization, innovation system construction, technology research and development, big data construction, etc.[4]. In 2022, the "Opinions on Promoting the Implementation of National Culture Digitization Strategy" was officially issued by the two State Offices [5]. A chapter on "Culture" was deployed in the Report of the Party's 20th National Congress that clearly proposed the implementation of "National Culture Digitization" strategy and strategy driven by major projects in the cultural industry. Since then, China has ushered a new track for the development in cultural digitalization. The 14th Five-Year Plan for National Economic and Social Development and the Outline of 2035 vision goals clearly put forward that by the end of the 14th Five-Year Plan period, China will build a multidimensional cultural service system featuring a complete network infrastructure, integration of online and offline services, and full coverage of cultural services [6]. It further encompasses the acceleration of the development of new cultural enterprises, innovation of cultural business forms and updating the cultural consumption patterns [7]. Also, it focused the development and strengthening of the online audio-visual platform, digital publishing, creativity and entertainment, and online performance. This initiative can improve the level of public culture digitization, accelerate the digital layout of the cultural industry and promote science and technology to empower the cultural industry, in order to ensure a cultural big data system of quick link, key integration and comprehensive sharing by 2035 [8].

In the last decade, the Chinese culture digitalization has comprehensively evolved and developed, and new formats, new scenes, new ideas and new models based on the innovations in digital technology have been continuously integrated into cultural life, cultural production and cultural ecology with fruitful development [9]. “Implementing National Culture Digitization Strategy” was a prodigious decision concerning the strategic arrangement made by the Party Central Committee to grasp the direction of digitization, intellectualization and network development. It also enabled an effective approach to promote the mutual interaction, mutual promotion and coordinated development of cultural industries, cultural undertakings and cultural forms of business [10]. The ‘data’ is an important factor in the production and development infrastructure, which has become an important focus of the Chinese national policies. With the evolution of the unstoppable and irreversible “digital age”, the “digital survival” is considered to be an important reality faced by the human race [11]. There is a need for increasing the participation and contribution of the digitization of culture. The “Implementing the Strategy of Digitization of Cultural Industry” is a strategic plan made by Communist Party of China based on the reality that the digital information industry can promote the rapid development of the society and economy. It is also a strategic choice to realize the 2035 vision target plan and the great rejuvenation of the Chinese nation.

2.2 Connotation of the cultural digitization

2.2.1 Connotation of digitalization

With the development of the Internet, digitization and informationization go hand in hand. The essence of the digitization process is the process of converting real things from digital equipment to analog signals for processing, storage, transmission or processing of binary code; introducing the computer to uniformly process into numbers or data through model converter; and then re-constructing these numbers or data into appropriate digital models. Digitization emphasizes the refinement and accuracy, truth and science, norms and standards, efficiency and benefits, and global and systematic norms. Digitization also greatly improves the product quality, shares information resources, and stimulates people’s pursuit of excellence in response to the product efficiency. Application of the digitization methods promote the advancement of science and technology and foster an improvement in the management level of all social strata. In the information age, the digitization of information is increasingly valued by researchers. All aspects of the people's daily life can be integrated using digital technology and can be “digitized”. As a result, the concepts of “digital economy”, “digital humanity” and “digital survival” continue to prevail. The application of digital technology has incubated a large number of new industrial forms, and the creation of new value by digital technology has promoted the emergence of new industrial models. Enterprises with new models recognized by the market have rapidly gathered into new social industries. “New business forms - new models - new industries” is the basic logical path for the cultural digitalization to promote the industrial development [12].

2.2.2 Connotation of cultural digitization

Industry + Digitalization present the implication of digitalization in the four links of the economic society, *viz.*, production, distribution, exchange and consumption. Digitalization profoundly affects the different forms of the industrial organization, efficiency of the industrial output, and structure of the human society and economy. It has huge subversive innovation ability for traditional industries and creates a grand economic size. "Cultural digitization" is a vivid practice of "digitalization" in the cultural industry. It refers to the use of modern video and audio coding, intelligent algorithm, digital copyright, network communication and other information digital technologies to store, create, transform, spread, consume and trade the existing various "non-digitalized" cultural resources to form a new "digital" culture. It can realize the upgradation of the cultural content, cultural transmission and cultural consumption mode, and assists the innovation of the existing cultural ecological system for the social and economic benefits. Apart from enhancing the process of cultural digital application in the process of digital display and integration of cultural resources, it also the process of amalgamation of national cultural big data system by using digital technology to serve the national cultural production, consumption and transmission channels [13]. Cultural digitization is a kind of sharing and service mode with interactivity, experience and innovation. Therefore, cultural digitization is not a single technical transformation of "using numbers to give wisdom", but is an iteration and innovation of the integrated digital cultural system.

The development of the digitization of Chinese culture has three stages, *viz.*, the stage of production and consumption of cultural content in the pure Internet mode; the two-way integration of traditional culture and the Internet; and the offline and online integration and application of highly developed digital technology. Currently, people look forward for the digitalization of the whole process of cultural "creation, production, marketing, management and service", so as to enjoy the cultural feast of excellence and get the ultimate spiritual pleasure. Cultural digitization is a stage and process of development in which technology, culture, humanities and economy constantly integrate, evolve and develop. From the technical aspect, cultural digitization should include the following contents: technology platform of hardware facilities (VR/AR, AI, spatial holographic projection, film and television digital shooting, cloud service platform for data processing, etc.), software facilities of file content creation and information processing (AI audio and video production software, 3D special effects production software, intelligent graphics generation software facilities, etc.), information flow and generated content of cultural digitalization, and human digital culture consumption pattern. With the development of cultural digitalization, it is moving forward to the digital and intellectual stage of culture, and the stage of cultural content digitalization is moving forward to the stage of cultural industry digitalization [14]. The hyper-realistic function of digital culture has given rise to a digital cultural virtual person and immersive virtual environment. Cultural digitization has fostered the cross-border integration of cultural travel and literature, and enabled them to be used in the related industries of society and economy, thereby increasingly forming a national cultural big data system with three-dimensional coverage, logical correlation, key integration and comprehensive sharing.

2.3 Value function of cultural digitization

The world is vibrant because of its interconnection, and the value of data is available and deployed around the corner and around the globe. Data factors provide an important multiplier effect on the capital, labor, technology, land and other production factors. In the digital information age, big data has become a new driving force for high-quality social leapfrog development. In the era of digital economy, data is an important factor of production, as well as, a basic strategic resource of the country [15]. The essence of cultural digitization is to correlate the data in the cultural field and continuously transform the related data containing the knowledge and wisdom of artists into tradable, traceable and quantifiable cultural assets. The core driving force of cultural digitalization is to promote the effective transformation and continuous development of the value of cultural assets and constantly activate the consumption power and source vitality of the cultural resources. The emergence of digital culture has rapidly changed the commercial space of retail, tourist attractions, mass consumer goods and other industries, and prompted significant changes in the service, marketing and production mode of many industries. Presently, life scenes such as e-commerce, live streaming, virtual human shopping guide, virtual human hosting and immersive visiting experience have become very common. The deep integration of digital culture and many industries continues to deepen and promote the iterative upgradation of many industries. Currently, the people's demand for digital culture shows a "blowout" trend. Furthermore, digital culture has become an effective way to meet people's high-quality spiritual and cultural needs. Cultural digitization enables the real economy and has become an important engine to promote the optimization and upgradation of the social and economic industrial structure. Cultural digitization is an internal requirement in promoting the high-quality development of the cultural industry. The national strategy of cultural digitization profoundly changes the overall structure, output value and industrial form of cultural industry. In addition, it influences and promotes the development of the cultural industry and social related industries. Thus, cultural digitization is an important starting point for China to achieve the goal of building a strong culture in 2035, and it is also a systematic project to achieve digital cultural governance under the leadership of the CPC [16].

2.4 Implementation routes of cultural digitization

The practice of cultural digital production, consumption and communication has been carried out in China for more than decades. In the digital strategy of China's cultural industry, it has been clearly proposed to build a "cultural big data service system". In this system, it is necessary to build a "national cultural private network" relying on China's radio and television network, and strengthen the construction of "national cultural big data cloud platform" relying on the national radio and television network. Furthermore, it also strengthens the construction of China's "digital cultural production line", promotes the integration of information technology and cultural elements, effectively integrates cultural industries and cultural undertakings, and fosters a series of new social and cultural business forms in China. The overall structure of the national cultural "big data system" can be summarized as two sides of supply and demand + four terminals [2], where the two sides of supply and demand refers to the demand and supply side of cultural digitalization, and the data generated in the consumption link of the digital culture field is called the demand side of digital culture. The data sources that people randomly pick from the national cultural big data system database is called the Supply side. The ‘four terminals’ include the resource terminal, production terminal, consumption terminal and cloud terminal of cultural digitalization. Libraries, museums, newspapers and other sections and units that keep or collect cultural resources form the “resource terminal” of cultural digitization. All the institutions that collect process and produce data for cultural resources make the “production terminal”. The institutions or places of digital culture consumption form the “consumption terminal”. Finally, the cultural data service platform or service center that serves the release, transmission, storage and transaction of cultural digitalization is called the “cloud terminal”. For example, cloud service platform and national cultural private network are both available via the “cloud”. In fact, institutions as public cultural resources can sometimes be regarded as both the “resource terminal” and the “production end” of cultural digitalization. “Cloud” is the link connecting the other “three terminals”. “Cloud” can realize the “interconnection” of data in various industries and fields and gathers rich and massive big data resources, which is the key for the big data to play the “magic” role.

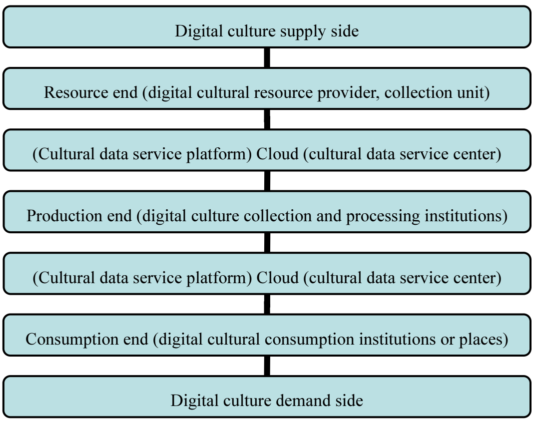


Figure 1. National cultural digitization framework system

Interconnection is the key for the successful implementation of the national cultural digitization strategy [17]. Chinese culture has a long history of more than 5,000 years. The cultural attributes and the elements of each unit, department or platform are the “gold hills”, but they still remain as the isolated islands of information. The major hurdle is faced in the extraction of cultural symbols and elements with inheritance value and to make overall use of the existing or under construction cultural achievements to create data with unified standards, scientific classification and labeling standards. Subsequently, under the support of “computing power” and “algorithm” can be correlated to the large database of Chinese culture and transformed into cultural production factors. This is the goal of implementing the national cultural digitization strategy.

3 Relationship between cultural digitization and library development

Digital cultural resources have become an extremely important factor of production under the condition of digital information. Cultural digitalization is a great historical process of the development of today's society. Cultural digitization converts the related works, knowledge and wisdom of cultural producers into quantifiable, traceable and tradable digital assets. It promotes the deep-level processing and service of cultural resources, extends the supply chain of cultural digital resources, and promotes the integration and development of different sectors. Cultural digitization strategy penetrates the whole process of cultural construction by promoting the upgradation and transformation of cultural resources, improving the digital service level of cultural resources, and stimulating the new vitality of cultural resources. Thus, it is an inspiration to promote the creative development of the national culture. Cultural digitization strategy is a great opportunity to realize the great rejuvenation of the Chinese nation. China has experienced the development process of "digitalization of cultural resources - digital application of cultural industry - operation of national cultural data system", and will soon adopt the new business model of integrated development of information technology and culture in a compound three-dimensional way. The "meta-universe" full intelligent life style of symbiosis of the virtual and real will be soon witnessed. Development of cultural digitization also truly presents the development of the new 'digital economy' from scratch. It is an inevitable result of modern information technology to promote the development of productive forces, and the inevitable result of cultural digital technology to promote the integration and development of other industries. In today's information age, mankind has gradually entered the era of "digitalism" or "intellectualism". The main contradiction of society is the "digital divide" between the demand for information and cultural digital resources and the lack of data in the process of social material production. The process of cultural digitization is a process in which human beings form a systematic and efficient cultural digitization system in the production, consumption and dissemination of cultural resources. It is a process in which human beings actively consider shortening the cultural digital gap. In order to stand at the forefront of the global information development, relevant cultural industries must fully adopt the existing facilities and networks, actively utilize the most advanced digital technology for resource integration, and promote cultural creative transformation and innovative development. This is not only a new growth point of the Chinese economy, but also a new level of the development of human culture.

Library is a cultural building with rich documents, a cultural market with rapid information circulation, and a cultural island with flexible spatial functions. The library itself is a culture with profound connotation, which is developed by the progressive inheritance of the past, the present and the future. It stands by the fusion and reflection of things (the library), people (readers and librarians) and gods (the wisdom and spirit carried by the collection of documents). With a development of the society, there is an ever-increasing importance of the library in the people's mind. There are several proverbs in this regard, including "Library is the door to knowledge", "library is a university without walls", "library is not only the organ of cultural crystallization, but also the organ of cultural propaganda", "one more library, one less prison". As a social and cultural institution that collects, keeps and arranges books, the library's mission is to provide the people with all kinds of published text, images and audio information freely and without reservation. This helps to create a good environment for the people to enjoy reading, improve their study and conduct scientific research. The library naturally bears the sacred mission of providing basic conditions for scientific research, independent decision-making, lifelong learning and cultural development of individuals or social groups, and naturally becomes a "growing organism". From ancient library to modern library, from modern paper library to modern digital library, the emergence of modern electronic library and network digital library is closely related to computer network technology, communication technology, and science and information technology. In the age of information data, people have higher requirements of reading in the library for "open environment, convenient service and efficient resource sharing". For decades, libraries have adhered to the service concept of specialization, individuation and digitalization by continuously building multifunctional electronic reading room, carrying out online exhibition activities, storing a large number of documents and video materials, and constantly strengthening the construction of information resources. In the digital age, cultural space will enter the new mode of "third space" from the compound platform, and "virtual space" will become the "door" through the real physical world. Reading is no longer a process of limited knowledge acquisition, but a process of "tasting culture, edifying sentiment and enlightening mind". People are increasingly demanding the immersive intelligent cultural space of the "condensed visual elements integrated with scientific and technological features, complex and multiple functions, and conveyed humanistic atmosphere" in libraries. At present, the construction of Shanghai Lingang Digital Science and Technology Library with digital immersion experience as the core is under way. By then, the library will provide readers with complete scientific and technological information and full sensory interactive reading experience with smooth international cooperation, which will be a model and benchmark for the construction of new digital libraries.

In the age of digital revolution, all cultural industries are facing great survival challenges and great new opportunities for development. In the new era, in order to fulfill the sacred mission of leading and promoting the reading of the whole people, there is a need to adapt to the new development and changes, open a new situation in the changing situation, and nurture new opportunities in the crisis. Libraries should deepen their understanding of the cultural digitalization strategy, and firmly grasp the historical opportunity of the new round of information revolution and industrial transformation. Further, they should strengthen the digital construction, actively expand the "Internet plus" and "smart plus", apply mobile Internet, big data, Internet of Things, block-chain and other technologies, and empower the library supply chain. Enhancing the library's response, perception, connection and operation capabilities in the construction of "national cultural big data service system" can provide a more and better efficient resource allocation. The implementation of cultural digitization strategy can effectively assist in the promotion of the transformation and upgrading development of libraries. Additionally, it provides a solid channel and assurance for the high-quality development of libraries, and paves a path and space for the development of libraries. The implementation of the national cultural digitization strategy can greatly promote the digital transformation and service innovation of libraries [18]. In view of the implementation plan, specific tasks and strategic objectives of "culture digitization", the library has ushered in an unprecedented significant development opportunity. The library must keep up with the present issues, uphold the innovative mentality, and maintain the forging attitude. In addition, libraries should constantly strengthen the digital construction by using advanced digital technology, and accumulate and transfer the useful and novel information resources. Thus it should "run fast with big steps when it is certain and practice with small steps when it is uncertain". There is a need to find a way to transform and develop this cultural digitization strategic projects that contain an infinite wealth and development space.

4 Development path of library under the background of cultural digitization strategy

4.1 Enabling technology and Upgrading the digital library technology application

With the rapid development in the network, communication and computer technology, "digital library" - a member of the digital family is known as one of the main information industries in the 21st century. The library integrates different digital technologies like information compression, optical character recognition and digital scanning to establish a distributed network resource database structure model. In the construction of cultural digitization, the basic requirement is to digitize all the existing paper documents and convert such documentation into electronic documents which is recognizable, retrievable, understandable and queried by computers. Subsequently, it can be applied to library database collection, storage and processing. Most libraries have employed advanced digital management system in the past, and purchased professional digital reading terminal equipment, including electronic book borrowing machine, self-service book borrowing and returning equipment, and digital resource collection equipment. However, these expensive digital management systems have not been effectively used due to the lack of visualization system, complete information digitization supporting equipment and related skilled information technology. This seriously hinders the depth and development of library digital construction. No matter how strong the construction of library information digitization resources or the construction of library think tank resources are to realize digitized, humanized and intelligent literature retrieval service, it is necessary to rely on advanced hardware equipment to upgrade a library's digital technology application system. This includes display screen, switch, network adapter, database management software, general or special application software and other software facilities. The library should develop and establish a corresponding digital application, management and service platform according to its own library management. Libraries should vigorously introduce system software, build a service platform by using data analysis system and visualization system, use AI, big data, cloud computing and artificial intelligence to build a library system management. It should be followed with a service system through resource sharing, monitoring and intelligent scheduling, and creating a digital library with "data-based management". There is a need to improve the library's ability of information mining, preservation, efficiently transmission and sharing of knowledge. This can help to realize the multi-content and high-quality cultural service supply of the library and meet people's spiritual and cultural needs at a higher level.

Cultural digitization is a powerful means to promote the innovative and integrated development of social resources. As the distribution center of information knowledge and human civilization, it is necessary to enhance the construction of library digitization. According to the requirements of the construction of "National Cultural Big Data Service System", the library should comprehensively sort out all kinds of information resources according to the principle of "logical correlation and physical distribution". Subsequently, it should scientifically classify them according to the standard marks of national cultural resources, and complete the collection, storage, distribution, transmission and processing of all library materials and data in the large system of national cultural private network. The establishment of library database in accordance with ISLI (International Standard Association Identifier for Information and Literature), initiated by China, facilitates the formation of Chinese cultural resources database. Based on the innovative and iterative development of digital technology, libraries should rely on the national cultural private network, build a data service platform, provide digital culture search and query, matching transactions, identity analysis and other services. This can assist in improving the digital service and creating a batch of first-class digital literature resources and culture, thereby realizing the continuous appreciation of the collection resources. As a cultural highland, libraries should strengthen the comprehensive and collaborative application of information technology, such as mobile Internet and communication technology. They should strive to use new information technologies such as artificial intelligence technology to find a wide and deep space for multi-type and multi-level digital application, and form a mechanism for digital innovation technology to augment the cultural integration and development. Also, they should form a driving force for the digital culture production, consumption, and sustainable innovation and development. To upgrade the application of digital technology in libraries, there is a need to strengthen the integrated application of 3D engine, virtual reality engine, digital human engine and other technologies. These can strengthen the integration of library information resources, improve the function of guidance, access and service of libraries, and give full play to the resource advantages of libraries in the ecological chain of "cultural communication".

4.2 Integration and innovation to consolidate the new library digital infrastructure

In recent years, new forms of digital culture such as webcast, short video and virtual environment have increasingly become people's way of life. Under the pressure of innovative information integration technologies such as AI, AR/VR/MR and FNT, information digitization has profoundly affected the whole process of human life in multiple fields. Digital empowerment of cultural heritage has led to a multi-dimensional immersive continuous experience [19], and the deep integration of information digital technology and culture will reshape the new form of social development. Science and technology are the primary productive forces, while network information technology is the key means of the digital construction of a library. Libraries must conform to the developing trend of the digital age, and thus develop new level of digital infrastructure on the premise of ensuring continued stability, scalability and compatibility of various library digital platforms. China's new digital infrastructure includes VR, AR, MR, virtual reality technology, Internet of Things, industrial Internet, artificial intelligence, integrated computing and other infrastructure [20], which is an important basis for the national digital construction. AR/VR/MR fosters immersive cultural experience, while the Internet of Things technology realizes the real-time simulation experience of cultural environment. The upgrade of artificial intelligence realizes the intelligent cultural experience. The application of digital technology reconstructs the relationship between "people, things and fields", and promotes the iterative upgrade of cultural consumption. Integrated information technology provides feasible strategies for the utilization, dissemination and protection of library collection cultural knowledge.

The new digital infrastructure can create tactile, rendering and immersive cultural products that can easily convert the complex data back and forth to realize the integration of digital resources. It can enable in-depth and high-quality development of cultural products to create a new environment for data circulation, and build a cultural lending system with supply and demand matching and efficient circulation. At present, audio-visual resources such as video and audio in a library only provides a two-dimensional display mode. If three-dimensional virtual reality literature can be generated with the help of virtual reality technology to make these precious literature resources "alive", the preservation and utilization of resources can be better realized. The library should adapt to the development trend of digital technology, actively use 3D digital, VR, AR and MR to accurately locate the location of books. Use of artificial intelligence and cloud computing should be enhanced to strengthen the underlying algorithm of the library. The block-chain, artificial intelligence, digital twin and other technologies can be employed to enable the contents, images and relics of library books to be presented in a multi-dimensional and three-dimensional way with realistic pictures. Finally, a "meta-universe" library can be created that integrates the virtuality and reality. There is a need to employ new virtual digital infrastructure to strengthen the collation, organization and indexing of library information, build an orderly and standardized digital information base, convert electronic books, image materials and information materials into 3D format, build a three-dimensional data information management system, and complete the basic work of digitized and digitized library literature information. In addition, it can realize the digital text description of library literature information, realize the upgrade and transformation of traditional library resource information and form a new information library. Alternatively, the "human-computer interaction" information technology can be used to extract the information from cataloging, circulation, classification, and collection, and digital processing can be enhanced by three-dimensional reconstruction, digital twinning, Internet of Things and other technologies. Readers can move into or join the virtual space with virtual identities, thereby realizing the dialogue scenes between readers and characters in the book across time and space, and allowing readers to "personally visit" the world in the book. Meeting the "He, she, it" in the book forms a vivid and flexible intelligent interactive interface for the reader. It realizes the readers' intelligent operation and three-dimensional visual desire in the link of literature retrieval and information browsing. Moreover, it realizes the readers' dream of convenient operation of the super-space-time simulation scene, so that the reader enjoys multi-dimensional dynamic and materialized reading and greatly reduces the distance of readers reaching "knowledge". Furthermore, it enhances the appeal, attraction and communication power of library collection resources, for the protection of Chinese culture, cultural inheritance, cultural transmission to fly high with wings.

4.3 Expand applications and improve library digital services

At present, the new forms of digital culture consumption show the characteristics and trends of "enhancing the senses, emergence of new consumer products, and creating chained wealth", and new service industries such as "related intellectual innovation" keep emerging. Libraries should actively link the scattered resources of different departments to connect the databases in different fields, and combine audio, pictures, text and other data in different forms. They should vigorously promote the timely use of library related resources on the cloud, and promote the deep mining and wide dissemination of collection resources. To make full use of digital resources and establish green ecology of digital resources, libraries should try their best to use digital technology to strengthen digital communication of all paper resources in their collections through text exhibition, picture display, audio and video playback. This can help to realize permanent communication and preservation of library digital resources that are simultaneously displayed offline and online. By combining online games, online animation and webcasting with the creation of new library digital industries, the library digital resources can be widely disseminated through terminals, so that people can conveniently enjoy the excellent cultural achievements of the collection without time and space restrictions. This leads to an improvement in the level of library digital reading and communication, and promotes the structural reform of library digital supply side. There is a need to promote the digital integration of library collection resources in all forms, processes and domains. This can be achieved by building the library digital resource database and building immersive new scene of library cultural experience. Improving the library computing service system and enhancing the supply capacity of library digital culture can lead to development of the universality and coverage of library digital services. Also there is a need to improve the digital service level of library resources, and realize people's desire for the digital culture of fast sharing and rendered experiences in the library.

Integrated technology can be adopted to build a library "cloud" resource data system, which can enrich the means of protecting library collection resources, actively link or associate with other data sources or cultural industry entities, and lay a solid foundation for further mining the value of library data resources. At present, many libraries are embarking on the "digitalization" project, but their digital development is only reflected in the network communication and sharing of the collection materials such as video and text, rather than "transferring the traditional content production and business development positions to the digital platform to activate the digital momentum of the library" [21]. In the process of digital construction, the library can actively recombine, deconstruct and create new 3D/4D works that are "constantly surprised and never get used to" by reorganizing, deconstructing and creating the document resource elements scattered around the corners of the library. This includes the text data and the image data, historical biography, text interpretation and graphic design of a certain theme (such as - A Dream of Red Mansions). Subsequently, it can cooperate with related cultural enterprises to create a new cultural consumption industry with wealth index. The library should take "serving cultural communication" as its core value and promote the development of new forms of business such as "immersive reading, intelligent library and meta-cosmic library" with its own cultural elements and digital elements of information. Also, it should promote the internal and external interaction and transformation of collection resources, strengthen the cooperation with cultural enterprises, and expand new scenes of consumption of digital cultural exhibitions with different themes. To create a batch of first-class digital culture consumption service products, the library should truly build into a "regional network center, nationwide reading education center, information and a literature resource service center". These steps can solely promote the library's excellent digital culture to the world in a deeper, more extensive and wider scope, and illuminate the development road of Chinese culture digitalization.

5 Conclusion

With the rapid development and strong drive of digital technology, the digital development road of Chinese culture is moving towards an important, efficient and intelligent direction. As one of the strategic positions of cultural power, the library should keep pace with the changing times, and focus on the key tasks and long-term goals of cultural digitalization strategy. There is a need to increase the digitization process of collection resources, increase the formation of library digitalization information flow, content flow and service flow. This will assist in highlighting the unique resource advantages and professional contributions of library in the construction of system of national cultural big data. The library should actively face up to the new situations, new technology and new problems in the process of digital development. It should strive to imbibe the technology empowerment, integration and innovation ability and application ability of the library's digital construction. This can provide more practical reference value for the digital development of Chinese culture, which is a vital expectation and requirement of the national cultural digitalization strategy.

References

[1] Zhang Manzi, Yang Shujun. The development of digital culture industry press the "acceleration key" [N]. Economic Information Daily, 2022-06-23 (006). http://dx.doi.org/10.28419/n.cnki.njjck.2022.002935
[2] Gao Shusheng. National Culture Digitization Strategy: Technical route and central link [J]. People's Forum • Academic frontier, 2022 (12): 14-19. http://dx.doi.org/10.16619/j.cnki.rmltxsqy.2022.23.002
[3] Tang Lin. Historical logic, Theoretical logic and practical Approach of digital transformation of cultural industry in 5G era [J]. Journal of Nanning Normal University (Philosophy and Social Sciences), 2021 (9): 58-67. http://dx.doi.org/10.16601/j.cnki.issn2096-7349.2021.05.006
[4] Fan Zhou. Thinking on the development direction and construction of National Cultural Parks under the background of cultural digitization Strategy [J]. People's Forum • Academic Frontier, 2022 (12): 48-54. http://dx.doi.org/10.16619/j.cnki.rmltxsqy.2022.23.006
[5] General office of the central committee of the communist party of China, the State Council general office. The opinions about digital strategy to promote the implementation of the national culture [EB]. [2022-5-22]. http://www.gov.cn/zhengce/2022-05/22/content_5691759.htm.
[6] Zou Tongqian. Current situation, problems and Countermeasures of Chinese Culture digitization [J]. People's Forum • Academic frontier, 2022 (23): 56-65. http://dx.doi.org/10.16619/j.cnki.rmltxsqy.2022.23.007
[7] Zhu Xudong. Exploration and practice of cultural digital industry development [J]. Smart China, 2022 (6): 52-53.
[8] Han Han. The Historical Evolution and Practice Path of the Communist Party of China Leading Culture Construction [J]. Journal of Shanxi University (Philosophy and Social Sciences Edition), 2021 (3): 12-20. http://dx.doi.org/10.13451/j.cnki.shanxi.univ(phil.soc.).2021.03.002
[9] Xie Xuefang, Zang Zhipeng. Accelerating the National culture digitization strategy [N]. Chinese Journal of Social Science, 2022-12-22 (001). http://dx.doi.org/10.28131/n.cnki.ncshk.2022.005674
[10] Zhou Linxing, Cui Yunping. Analysis on the construction path of archival culture under the National culture Digitization strategy [J]. Archives Science Communication, 2023 (2): 10-17. http://dx.doi.org/10.16113/j.cnki.daxtx.2023.02.005
[11] Liu Hanyu. Digital Thinking: Guide for Digital Transformation of Traditional Enterprises [M]. Beijing: China Machine Press, 2021: 29-30.
[12] Yu Liangzao, Meng Rui. The "Strong" of Socialist cultural power [J]. Henan Social Sciences, 2021 (3): 109-117.
[13] Gu Jiang. Development strategy innovation of digital cultural industry from the perspective of cultural power [J]. Journal of Shanghai Jiao Tong University (Philosophy and Social Sciences Edition), 2022 (4) : 12-22. http://dx.doi.org/10.13806/j.cnki.issn1008-7095.2022.04.002
[14] Li Yiruo. Exploration on high-quality development path of rural cultural industry under Culture Digitization strategy [J]. Wealth Today, 2023 (1): 26-28.
[15] Zhang Ying, Zhang Xiaobin. New thinking at the bottom of cultural digitization [J]. Culture and Industry, 2023 (1): 30-32. http://dx.doi.org/10.16045/j.cnki.rti.2023.01.003
[16] Fan Zhou, Sun Wei. Research on the Development process and Path of National Culture Digitization Strategy [J]. Journal of Central China Normal University (Humanities and Social Sciences Edition), 2023 (1): 70-76. http://dx.doi.org/10.19992/j.cnki.1000-2456.2023.01.007
[17] Guo Quanzhong, Wu Xiaoxin. Key and key points of successful implementation of National Culture Digitization Strategy [J]. China Development Observation, 2022 (10): 27-31.

[18]Xu Yi, Qian Zhiyong. Library Innovation Practice facing the National Culture Digitization Strategy -- Taking the digital protection and utilization of rare endangered literature on Zhang Jian Studies as an example [J]. New Century Library, 2022 (11): 21-29. http://dx.doi.org/10.16810/j.cnki.1672-514X.2022.11.003
[19] Zhou Jianxin. Digitization of Chinese excellent traditional culture: Logical approach and practical innovation [J]. Cultural Industry Research, 2022 (10): 82-88. http://dx.doi.org/10.14180/j.cnki.1004-0544.2022.10.010
[20] Research Group of People's Think Tank. Digital transformation enables central enterprises to develop with high quality – FAW's digital intelligence moment [J]. National Governance, 2021 (48): 3-9. http://dx.doi.org/10.16619/j.cnki.cn10-1264/d.2021.48.001
[21] Li Yuchen, Li Yan. Key issues of cultural industry digitization [J]. People's Livelihood Public Opinion, 2022 (24): 88-90.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230186

Does the Development of the Digital Economy Benefit Urban Carbon Reduction? Evidence from the Yangtze River Delta Region

Jungang Shao[1] and Qing Li
School of Shanghai Maritime University, Shanghai, China

Abstract. The Yangtze River Delta (YRD) is an experimental location of high-quality holistic growth in China and an important area for realizing the "dual carbon" goal. Utilizing panel data from 27 important municipalities in the YRD urban agglomeration from 2010 to 2020, the fixed-effect model is used to evaluate the stage of advancement in the digital economy on urban emission levels. As an intermediary variable, industrial structure modification was tested to see how the digital economy affected local dioxide emissions. The results show that: (1) After an array of robustness tests, the overall impact on both core variables is inverted U-shaped, facilitating and then constraining emissions; (2) As part of its function in cutting emissions, the digital economy also impacts the industrial framework, which has an indirect, nonlinear, U-shaped effect on the intensity of regional dioxide emissions; (3) Looking further, regions with high levels of government spending and investment in innovation are those where carbon emissions are most clearly affected by digitalization. There is still opportunity for evolution and optimization of the YRD region's energy consumption process. The study's findings expand our comprehension of the factors that influence high-quality community growth as well as how the digital economy works to support it.

Keywords. Digital economy; carbon emissions; Yangtze River Delta (YRD); industrial structure

1. Introduction

Carbon emissions and numerous connected environmental issues have received global attention. The YRD region stands out as a hub of intense economic development, high levels of openness, and strong innovation capacities in China. The concept of digital economy states that the full evaluation index of the sector is built from the two dimensions of “digital industrialization” and “industrial digitization” by CAICT (2019). Most studies about digitalization are focused at the scale of the province, and a few are measured at prefecture-level cities and below in a comprehensive index evaluation method [1]. The digitalization can be measured from several levels, for example, digital innovation capability, digital business emergence, digital infrastructure, and digital trade

[1] Corresponding Author: Jungang Shao, School of Shanghai Maritime University, Shanghai 201900, China. Email: jgshao@shmtu.edu.cn.

level [2]. Meanwhile, the methods used to calculate dioxide emissions mainly include IPCC method, life cycle method, mass balance method and actual measurement method. Digital economy has a multi-path and multi-dimensional comprehensive impact on carbon emissions. Also, it can change urban environments in three areas: production, life, and ecological space and examined the effect of digitalization on the transition to a greener urban environment [3]. From a different perspective, digitalization has greatly decreased the sum of emissions and has a nonlinear relationship with intensity of emissions, which shows as a U-shaped curve [4,5]. Furthermore, comparing with the labor efficiency improvement effect, reducing energy intensity can further promote reducing the emissions [1].

However, the possible margin of contribution of this research in comparison to the existing literature includes: First, most of the current academic study in this field is provincial or national, and relatively few studies are based on urban agglomerations, our study enriches relevant research experience. Second, to better understand how industrial structure may play a part in the nonlinear link of the benchmark result, this research also adds industrial structure as an intermediate variable. Third, we conduct heterogeneity analysis for investigating the variations in how the digitalization affects dioxide emissions under various policy contexts.

In summary, employing the data of 27 YRD's cities from 2010 to 2020, we attempt to construct a development of digitalization index system from the definition of digitalization, and use the fixed effect model to empirically analyses the effect mechanism, as well as the intermediary effect of industrial structure.

2. Research hypothesis

The following three factors may show how digitalization may affect carbon emissions: Digitalization has significantly boosted the prosperity of e-commerce market, given consumers access to a wider range of goods and services, greatly boosted demand, and allowed businesses to scale up production, which has increased usage of resources and carbon emissions [6]. Contrarily, it may also lead to the phenomenon of "Jevons Paradox" [7], that is, the efficiency of energy production will greatly improve the energy supply capacity, and in turn, the market competition of energy companies will intensify, leading to a decline in energy prices, promoting energy consumption.

According to the preceding analysis assumptions:

H1: The development of digitalization first increases and then reduces dioxide emissions, forming an inverted U-shaped curve.

Two factors may reflect how digitalization is affecting the manufacturing framework: (1) The digital economy can restructure the traditional energy sectors which makes it necessary to update or rebuild a cleaner and more complete manufacturing industry at the early stage of the development cycle. (2) The improvement of digitalization can hasten the transfer of information and production components between geographical areas, truly raising the bar for intelligent production while fostering the production of labor, land, technology, capital, etc. which can raise the proportion of the tertiary sector with relatively low emissions of carbon.

According to the preceding analysis assumptions:

H2: The digital economy indirectly affects regional dioxide emission by affecting the industrial structure.

3. Research design

3.1. Model Design

In formula (1), the subscripts i and t represent the city and the year respectively, the explained variable CE_{it}represents the carbon emission intensity of the i-th city in the t-year, and the explanatory variables DE_{it} and SDE_{it} represent the digital economy development comprehensive index of the i -th city in the t-year respectively the first-order and second-order items of , controls are a series of control variables, which $\mu_i$are city fixed effects, $\delta_t$time fixed effects, $\varepsilon_{it}$and random disturbance items.

$$CE_{it} = \alpha_0 + \alpha_1 DE_{it} + \alpha_2 SDE_{it} + \sum_{i=3}^{n} \alpha_i controls_{it} + \mu_i + \delta_t + \varepsilon_{it} \quad (1)$$

The intermediate effect test and intermediary effect model are built by referring to Wen's (2006) [8] research techniques in order to investigate the potential intermediary effect of digitization on the adjustment of industrial structure:

$$IS_{it} = \beta_0 + \beta_1 DE_{it} + \beta_2 SDE_{it} + \sum_{i=3}^{n} \beta_i controls_{it} + \mu_i + \delta_t + \varepsilon_{it} \quad (2)$$

$$CE_{it} = \gamma_0 + \gamma_1 DE_{it} + \gamma_2 SDE_{it} + \gamma_3 IS_{it} + \sum_{i=4}^{n} \gamma_i controls_{it} + \mu_i + \delta_t + \varepsilon_{it} \quad (3)$$

In formulas (2) and (3), IS_{it}it represents the i-th city's industrial structure in t-year as an intermediary variable.

3.2. Variable Description

Carbon emission intensity (CE) as the interpreted variable is measured by common fossils, this study employs data on the energy supply that is rather simple to get and is based on the methodology of Miao et al. (2022) [9] and uses relatively easily available energy supply data to calculate common fossil, which includes crude oil, kerosene, natural gas, etc. The following equation is adopted to calculate the intensity in prefecture-level cities and is based on large part on the carbon emission factor technique described in the IPCC Inventory Guidelines:

$$C_E^t = \sum_{i=1}^{n} C_{Direct,i}^t = \sum_{i=1}^{n} \sum_{j=1}^{m} [E_{ij}^t \times LCV_{ij}^t \times CC_{ij}^t \times COF_{ij}^t \times \frac{44}{12}] \quad (4)$$

Among them, C_E^t represents the total amount of carbon emissions; E_{ij}^t represents the j energy consumption of the i region in t-year; LCV_{ij}^t represents the average low calorific value of the energy; CC_{ij}^t represents the carbon content per unit calorific; COF_{ij}^t represents the oxidation rate of an energy source; $\frac{44}{12}$ represents the ratio of the molar mass of carbon dioxide to the molar mass of carbon element.

Comprehensive index of digital economy development (DE) and DE square (SDE) are explanatory variables. Considering the accessibility of data and the scientific nature of system design, this paper draws lessons from Xu et al. (2022)[10], Xie (2022) [11], Zhao et al. (2020) [12] for the construction of the digital economy index system , employing the percentage of urban regions' Internet-broadband users, the amount of tele-business conducted per capita, and the number of smartphone users per 100 people, and the digital inclusive financial index, the comprehensive index is obtained through the entropy weight method.

Industrial structure (IS) is the intermediary variable of this article, learning from the research of Gan et al. (2011) [13], it selects the ratio that the added value of the tertiary industry is divided to the secondary in prefecture-level cities.

The control variables are set for researching the function of digitalization in urban emission reduction more thoroughly, as follows: Government interference (GOV) is measured as the ratio that local general budget expenditure divides to regional GDP. Energy consumption structure (ECS) is calculated as the ratio of fossil fuel usage to total energy usage. Urbanization rate (UR) is the percentage of built-up area to urban area. Transportation capacity (ROAD) is the logarithm of actual road space per person. Technical level (SCIENCE) is the percentage of local science expenditure to GDP.

In this study, the data includes 27 prefecture-level cities of YRD are chosen from 2011 to 2020. The "China City Statistical Yearbook", "China Energy Statistical Yearbook" and other city-specific statistical yearbooks are the origination of data. When values are absent, linear interpolation is used to fill them. Table 1 displays the descriptive analysis of all variables considered.

Table 1 Variable descriptive statistics

Variables	ce	de	sde	is	gov	ecs	ur	road	science
N	270	270	270	270	270	270	270	270	270
Mean	0.01	0.174	0.038	1.011	0.142	0.114	0.123	2.568	0.007
Std.Dev	0.004	0.089	0.041	0.379	0.049	0.081	0.091	0.365	0.016
Min	0.004	0.061	0.004	0.269	0.075	0.041	0.015	1.396	0.001
Max	0.034	0.454	0.206	2.751	0.283	0.476	0.789	3.343	0.163

4. Empirical analysis

The F-test and Hausman test are utilized in selecting a suitable model in research. In the individual fixed effect model (column 1), The overall digital economy development index's regression coefficient is -0.021, which is statistically significant ($p<0.01$), but in the column 2, after controlling the individual effect and time effect, there is a positive correlation among main variables ($p<0.05$). It is likely that institutional factors will affect digitalization and emissions simultaneously, and those factors are affected by time larger, making some variables become endogenous variables, so this model considering the interaction items of temporal dummy variables can better exclude the influence of such unobservable factors. The coefficient of DE is positively associated with independent variable, while its quadratic coefficient demonstrates a significant negative relationship. This finding supports Hypothesis 1 outlined in this paper.

The coefficients of energy usage framework and government intervention level are significantly positive ($p<0.01$), demonstrating that as government spending grows, the coal's share of energy use rises. It may lead to the time lag of policy implementation, also government fiscal expenditure is not yet fully utilized. The negative coefficients obtained from the regression analysis of urbanization rate and transportation level are statistically significant ($p<0.01$), indicating that greater urbanization and easier access to roads have a positive effect on lowering emissions. The coefficient of technological development level shows that the increase in science expenditure may suppress carbon emission, but this effect is not significant in the current model.

Table 2. Model regression results

	(1)	(2)	(3)	(4)	(5)
	ce_gdp	**ce_gdp**	**ce_gdp**	**is**	**ce_gdp**
de	-0.021 ***	0.006 **	0.022 **	-1.407 *	0.024 ***
	(0.004)	(0.003)	(0.009)	(0.954)	(0.009)
sde			-0.026 *	4.231 ***	-0.032 **
			(0.014)	(1.497)	(0.014)
is					0.001 **
					(0.001)
ecs	0.024 ***	0.006 ***	0.007 ***	0.256	0.006 ***
	(0.002)	(0.002)	(0.002)	(0.174)	(0.002)
gov	0.033 ***	0.054 ***	0.056 ***	-0.441	0.056 ***
	(0.009)	(0.005)	(0.005)	(0.572)	(0.005)
ur	-0.008 ***	-0.004 ***	-0.004 ***	-0.116	-0.004 ***
	(0.003)	(0.002)	(0.002)	(0.171)	(0.002)
road	-0.006 ***	-0.001 ***	-0.002 ***	-0.128 **	-0.001 ***
	(0.001)	(0.000)	(0.000)	(0.050)	(0.000)
science	-0.048 ***	-0.003	- 0.001	2.773 ***	-0.003
	(0.015)	(0.009)	(0.009)	(0.994)	(0.009)
_cons	0.022 ***	0.010 ***	0.008 ***	1.240 ***	0.006 ***
	(0.002)	(0.001)	(0.002)	(0.185)	(0.002)
N	270.000	270.000	270.000	270.000	270.000
r2	0.594	0.881	0.883	0.666	0.885
r2_a	0.539	0.859	0.861	0.605	0.863
individual	control	control	control	control	control
time	not control	control	control	control	control

Standard errors in parentheses * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

In this paper, industrial structure is used as an intermediary variable to build models, in which column (4) in table 2 represents formula (2), and column (5) represents formula (3). Column (4) shows that there is a U-shaped nonlinear link between digitalization and manufacturing framework. The digitalization leads to the updating and iterating of industrial equipment, which increases the consumption and added value of the secondary industry. Industries in the area with high energy usage and pollution have been influenced in more recent times by policy implementation and public pressure. The results in column (5) make it clear that there is a positive correlation between the first-order coefficient of DE and the intensity of emissions, while the second-order term shows a negative correlation ($p<0.05$), showing that the nonlinear, U-shape curve between main variables, in which the coefficient of IS is 0.001 ($p<0.05$). After eliminating intermediary variables in column (3), the absolute value of item coefficients for first-order and second-order of DE increased, while significance level also showed an increase when compared to its counterpart in column (5). This provides evidence that the digitalization has an indirect impact on carbon emissions through the intermediary variable of manufacturing framework, thereby verifying hypothesis 2.

4.1. Robust Test

The robust test includes three parts, firstly, considering the possible impact of extreme values of results, we replace the dependent variable as the original winsorized by 1%. Secondly, we change the method of calculating the the comprehensive index of DE from the entropy weight method to the PCA [14]. Thirdly, considering that policy-related control variables may have endogeneity, we use the level of government interference and the level of technological development as two tools with a one-period lag variable [15].

Table 3 Robustness analysis.

	(1)	(2)	(3)
de	0.021 ***	0.233 ***	0.034 *
	(0.007)	(0.075)	(0.022)
sde	-0.026 **	-0.155 ***	-0.039 ***
	(0.012)	(0.054)	(0.013)
_cons	0.008 ***	-0.073 ***	0.021 ***
	(0.001)	(0.025)	(0.003)
N	270.000	270.000	243.000
r2	0.906	0.889	0.567

The outcomes of the above three robust tests are shown in Table 3. In comparison to the benchmark regression findings presented in Table 2, it is evident that while there are differences observed in variable significance and coefficient magnitudes, the inverted U-shaped relationship still holds. Thus, this study's conclusion remains robust.

4.2. Heterogeneity analysis

Heterogeneity analysis is carried out from two perspectives. On the one hand, the sample is split into two equal portions by fiscal expenditure, that is measured by the percentage of area general fiscal expenditure in GDP, named high-level and low-level. Table 4 (1) and (2) display the findings. It may be concluded that the results are more significant in the cities with greater fiscal spending levels. On the other hand, this study separates urban innovation into two groups by the median level of technological advancement, citing Guo et al.'s approach from 2022 [16]. Column (3) and (4) of Table 4 display the regression findings. It is evident that the first-order and second-order terms of DE in cities with strong innovation input exceed the 5% significance level but fail to pass the significance test in cities with low innovation input.

Table 4 Heterogeneity analysis.

	Fiscal expenditure level		**Urban Innovation Level**	
	(1) **Low**	**(2)** **high**	**(3)** **Low**	**(4)** **high**
de	0.011	0.051 ***	-0.004	0.023 ***
	(0.008)	(0.016)	(0.018)	(0.007)
sde	-0.002	-0.081 ***	0.003	-0.024 **
	(0.010)	(0.027)	(0.033)	(0.009)
_cons	0.012 ***	0.003 ***	0.016 ***	0.008 ***
	(0.002)	(0.003)	(0.003)	(0.001)
N	135.000	135.000	135.000	135.000
r2	0.959	0.881	0.902	0.955
r2_a	0.945	0.842	0.869	0.939

5. CONCLUSION

From this study, we find the impact of digitalization on dioxide emissions shows an inverted U-shaped nonlinear pattern that first increase and then reduce. Meanwhile, within the time frame analyzed in this paper, manufacturing framework serves as a mediating factor. Also, this impact is more significant in regions with high fiscal expenditure and high innovation investment.

Considering the previous results, firstly, we propose to create a low-carbon industrial structure system in the YRD by a comprehensive plan, which will guide the upgrade of industrial structure and encourage to use greener energy, while simultaneously constraining the scale of high-pollution industries and high-carbon emission energy industries within the region. Secondly, to promote green development, it is imperative to enhance the low-carbon industrial system and the digital economy development standard system. To achieve this, we must foster digital innovation and promote economic growth that is both sustainable and environmentally friendly. Thirdly, the local policy on environment assessment is still in its early stages and has not yet completely tapped into the potential for energy saving and emission reduction, as well as the economic benefits of digitalization. To address these issues, it is important to increase urban science expenditures and enhance environmental monitoring capabilities and management flexibility.

The innovation of this paper is to study the relationship of digitalization and carbon emission intensity of urban agglomerations from a new research scale, to provide effective experience for digital emission reduction in different urban agglomerations and regions. Future researchers may benefit from these findings in their understanding of internal mechanisms among variables and in their evaluation and guidance of local governments' emission reduction initiatives. Also, the data used in this paper may not fully reflect the level of variables given the availability of data. Next, we can establish a more precise indicator system for digitalization and more comprehensive greenhouse gas data for calculating carbon emission indicators. The analysis of regional disparities study also focus mostly on economic factors. The next step is to investigate the driving mechanism from various angles considering the surroundings.

References

[1] Wang S, Yu DH. Carbon Emission Reduction Effect and Pathways of Digital Economy: An Empirical Investigation Based on the Carbon Emission Efficiency of China's Manufacturing[J/OL]. Studies in Science of Science:1-14[2023-05-21].https://doi.org/10.16192/j.cnki.1003-2053.20230516.002

[2] Li J L, Chen L I, Chen Y, et al. Digital economy, technological innovation, and green economic efficiency−empirical evidence from 277 cities in China [J].Managerial and decision economics, 2022(3):616-629

[3] Chang X, Su J, Yang Z. The Effect of Digital Economy on Urban Green Transformation—An Empirical Study Based on the YRD City Cluster in China. Sustainability. 2022; 14(21):13770. https://doi.org/10.3390/su142113770

[4] Ge Li-vu, Mo Long-iiong, Huang Nian-bing. Development of Digital Economy, Upgrading of Industrial Structure and Urban Carbon Emission[J]. Modern Finance and Economics-Journal of Tianjin University of Finance and Economics,2022,42(10):20-37.DOI:10.19559/j.cnki.12-1387.2022.10.002.

[5] Zhang Zhengyan, Li Yuxi. Research on the Impact of Digital Economy on Carbon Emissions in China,2022,43(05):146-154.DOI:10.16339/j.cnki.hdxbcjb.2022.05.019..

[6] Chen Xiaohong, Li Yangyang, Song Lijie,etc. Theoretical Framework and Research Prospect of Digital Economy[J].Journal of Management World,2022,38(02):208-224+13-16.DOI:10.19744/j.cnki.11-1235/f.2022.0020.

[7] Luo Liangzhong. Research on the Influence of Digital Economy on Energy Consumption[J]. Study and Practice,2022,No.460(06):44-53.DOI:10.19624/j.cnki.cn42-1005/c.2022.06.013.

[8] Wen Zhonglin, Zhang Lei,Hou Jietai. Mediated Moderator and Moderated Mediator[J]. Acta Psychologica Sinica,2006(03):448-452.

[9] Miao Lujun, Chen Jing, Fan Tianzheng, Lv Yanqin. The Impact of Digital Economy Development on Carbon Emission: A Panel Data Analysis of 278 Prefecture-Level Cities[J]. South China Finance,2022(02):45-57.

[10] Xu Weixiang, Zhou Jianping, Liu Chengjun. The Impact of Digital Economy on Urban Carbon Emissions: Based on The Analysis of Spatial Effects[J]. Geographical Research,2022,41(01):111-129.

[11] Xie Yunfei. The Effect and Mechanism of Digital Economy on Regional Carbon Emission Intensity[J]. Contemporary Economic Management, 2022,44(02):68-78.DOI:10.13253/j.cnki.ddjjgl.2022.02.008.

[12] Zhao Tao, Zhang Zhi, Liang Shangkun. Digital Economy, Entrepreneurship, and High-Quality Economic Development: Empirical Evidence from Urban China[J]. Journal of Management World,2020,36(10):65-76.DOI:10.19744/j.cnki.11-1235/f.2020.0154.

[13] Gan Chunhui, Zheng Ruo-gu, Yu Dianfan. An Empirical Study on the Effects of Industrial Structure on Economic Growth and Fluctuations in China [J]. Economic Research Journal, 2011,46(05):4-16+31.

[14] Jin Dianchen, Liu Shuai, Chen Xin. Digital Economy and Common Prosperity-An Empirical Test based on 276 Prefecture-level Cities [J]. Journal of Xinjiang Normal University (Edition of Philosophy and Social Sciences),2023,44(03):127-136.DOI:10.14100/j.cnki.65-1039/g4.20230227.001.

[15] Wang, LJ., Yang, PL., Ma, JJ. et al. Digital economy and industrial energy efficiency performance: evidence from the city of the YRD in China. Environ Sci Pollut Res 30, 30672–30691 (2023). https://doi.org/10.1007/s11356-022-24353-2

[16] Gou Feng, Yang Shang-guang, Ren Yi. The Digital Economy, Green Technology Innovation and Carbon Emissions－Empirical Evidence from Chinese City-level Data [J]. Journal of Shaanxi Normal University(Philosophy and Social Sciences Edition),2022,51(03):45-60.DOI:10.15983/j.cnki.sxss.2022.0507.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230187

Research on the Influence of Agricultural Products Brand Story on Consumers' Purchase Intention Under the Background of Social New Retail

Jia WEI [1]
School of Economics and Management, Lanzhou University of Technology, Lanzhou, China

Abstract. Under the background of the trend of social new retail, based on narrative transmission theory, this study collects 208 samples by questionnaire, integrating social media interaction and consumers' purchase intention, and analyzes the conditional process of brand stories affecting how to affect consumers' purchase intention by hierarchical regression method. The result shows that brand story has a positive impact on consumers' purchase intention, and this process is realized through the intermediary of narrative transmission. Social media interaction negatively regulates the relationship between brand story and consumers' purchase intentions. Finally, according to the research conclusion, it is suggested that enterprises and marketers pay attention to the role of brand story, make full use of various social media platforms, tell excellent brand stories, increase emotional resonance, and promote consumers' purchase intention.

Keywords. brand story, interactivity of social media, consumers' purchase intention

1. Introduction

Social New Retail is a new model derived from e-commerce platforms, which reinvents the elements of people, goods, market, and circle of friends, and derives a new business model with social circle as the core, and is committed to expanding brand content in the social circle. It is a mode for enterprises to open up online and offline, realizing seamless connection between online and offline scenes, enabling online sales to provide user experience of physical store and physical stores to connect online information and logistics. At the same time, we will realize the deep integration of social economy and new retail, and create a new retail ecosystem that constantly optimizes user experience and meets differentiated needs. How should agricultural products brands develop under the background of new social retail? As a big agricultural country, China's agricultural products market has changed from seller's market to buyer's market, and the competition among agricultural products has deepened to the brand level. Brand is an important resource for

[1]Corresponding Author: Jia WEI, School of Economics and Management, Lanzhou University of Technology, Lanzhou, China, 730050; E-mail: lutjiawei@126.com.

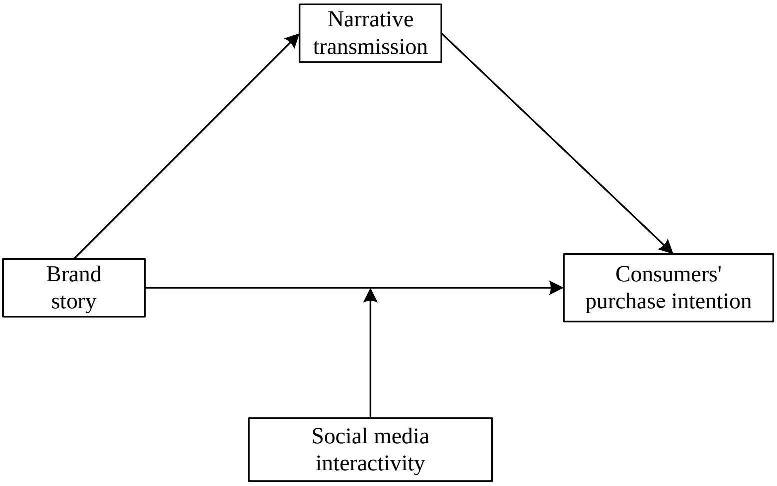


Figure 1. Theoretical model diagram

enterprises to participate in competition and an important business card for the country to show its image to the outside world. In addition, in the Internet age, the amount of information is overloaded, the channels of information collection are increasingly abundant, and the speed of communication has achieved a qualitative leap. Consumers have lost their focus and interest in traditional advertisements and empty and sensational slogans. Consumers' demand and behavior for agricultural products are more differentiated [1].

Compared with other products, the online marketing of agricultural products is more challenging, and it is very important to understand consumers' reaction to the brand promotion of agricultural products [2]. Related researches show that brand story can establish a connection between consumers and brands [3], and transmitting brand story can strengthen the connection, thus generating positive attitudes and higher purchase intention [4]. Therefore, agricultural products enterprises should be good at telling stories (telling story of farmers, landmark products, cultural tourism, and world heritage in combination with agricultural products) when providing products and services, and tell them with the greatest communication effect-moving people with emotion, touching people with reality, arousing the resonance of target users, feeling the connotation of brands, establishing brand loyalty of customers, conveying strong brand values, and online comments, sharing and recommendation, etc., which can make agricultural products enterprises establish a stable connection with consumers. Narrative forms enhance traffic and intentions, and immerse themselves in story, thus having a positive impact on consumers' purchasing intentions [3]. Narrative persuasion can take place in the context of consumer reviews, and the process is the same as hypothetically.

Based on this, this study constructs the relationship model between brand story and consumers' purchase intention under the background of new social retail, and explains its internal mechanism, so as to provide relevant suggestions for agricultural products enterprises and marketers under the background of new social retail, as shown Figure 1.

2. Theoretical Basis and Research Hypothesis

2.1. Brand Story and Consumers Purchase Intention

With the increasing power of consumers, people participate in brand story through social media, which attaches great importance in improving brand awareness, image, and loyalty. Storytelling is in every corner [5]. Therefore, marketers need to better understand the value of brand story [6]. The positive psychological persuasion process of brand story to consumers is explained with narrative transmission theory and narrative persuasion theory. Firstly, consumers are brought into the story to get an immersive experience and will change their attitudes and behaviors because of empathy, which is manifested in their story-consistent attitude and story-consistent intention. An attractive brand story can change consumers' positive attitude towards the brand and persuade consumers to purchase. The research shows that the influence of brand story with different attractiveness on consumers' brand attitude is different, and the influence of brand story with high-quality story content on brand attitude is more significant. Obviously, brand story has a positive impact on consumers' brand attitudes. Therefore, the following assumption can be made:

H1: Brand story will positively affect consumers' purchase intention

2.2. Mediating Function of Narrative Transmission

The foundation of telling a good brand story is the story itself. Excellent agricultural product brand story that conforms to enterprise values and beliefs is the premise of narrative transmission.

When individuals are "transported" in some way, they are completely immersed in the story, and the changes after returning to the real world are mainly reflected in the changes in thinking levels such as beliefs and attitudes. Marketing research also responds to this view, and holds that the key to improving consumers' pleasure lies in making consumers reach a focused psychological state, narrative transmission state, so that consumers can have positive emotions and involvement and enhance consumption experience. At present, there are also studies focusing on the follow-up influence of narrative transmission, such as repeated exposure, information search, memory, emotional change, social sharing behavior, etc., in order to maintain continuous interaction with consumers and further strengthen consumers' attitudes and behaviors. The stronger the relationship between consumers and brands, the stronger their willingness to spread online word of mouth. Enterprises pass on brand image and concept to consumers through brand story, cultivate consumers' brand emotion, and make them form preferences according to their cognition and perception of brands before purchasing goods or services.

H2: Narrative transmission mediates between agricultural brand story and consumers purchase intention

2.3. Interactive Regulation of Social Media

In the era of more and more active social media, establishing and maintaining a stable relationship between enterprises and consumers is the top priority of marketing [1]. The interaction between consumers and enterprises is faster, and positive interaction with consumers can achieve the purpose of promoting communication and purchase. As consumers, they also have more channel choices and greater initiative to spread product use and experience [7]. Consumers' access to content and information related to brand story in social media will interfere with consumers' views on brands and brand story, and the interference effect may be affected by consumers' social media interactivity [8]. The consumers' social media interactivity is measured from two dimensions: online comment, forwarding and sharing, or recommendation. Advertisements and information from brand will make consumers have positive associations with brand, which means that the difference of stimulation brought by brand story to consumers will be relatively small. Therefore, the following assumption can be made:

H3: Social media interactivity moderates the relationship between brand story and consumers' purchase intention, that is, social media interactivity weakens the influence of brand story on consumers' purchase intention.

3. Research Method

3.1. Variable Surveying

3.1.1. Brand Story

This study focuses on the brand story, mainly from the story of the event and characters of two aspects of the characteristics to measure the attitude of the characters and the authenticity of the event, adapted from the Escalas [9] research scale.

3.1.2. Narrative Transmission

Narrative transmission has the ability to connect consumers and brands [6]. The research results show that narrative transmission is positively correlated with advertising goodwill, and story trigger more favorable emotional reactions of audiences, and have a certain impact on audiences' willingness to share product information through word of mouth [10]. Narrative transmission can be measured by the Transportation Image Model (TIM) scale, which investigates two factors related to transmission degree: image ability and absorption tendency [11]. Image ability is defined as the ability of consumers to visually rehearse scenes through psychological images without imposing stimuli. Absorption tendency is defined as the reader's absorption of story content [11].

3.1.3. Social Media Interactivity

The measurement angle of social media interactivity is mainly based on two dimensions summarized by Ye qing (2013) and Su Qisheng (2015) according to the existing research and actual situation when conducting social media research: online comment, forwarding and sharing. The dimension of forwarding and sharing is measured with reference to the scale developed by Spannerworks (2007), and the dimension of participating in comment is measured with reference to the scale developed by Chen Shunlin (2007). This measurement dimension division and way have been tested by many scholars, which not only has been verified theoretically, but also conforms to the use of social media and the reality of our country.

3.1.4. Consumers' Purchase Intention

Purchase intention refers to the possibility of consumers' purchase intention specific products [12]. The measurement of purchase intention mainly refers to the measurement of consumers' purchase intention by Bauer [1], and the scale adopts the measurement method of one-dimensional and Likert five-level scale.

3.1.5. Control Variable

In order to avoid irrelevant variables interfering with the causal relationship between variables, the gender, age, educational attainment (EA), occupation, and disposable income/month (DIM) are selected as the control variables.

3.2. Samples and Data

Questionnaire survey is used to collect data. Use the questionnaire star to design the questionnaire and fill in the instructions, and send the link of the questionnaire through mobile internet tools such as WeChat and QQ. Necessary control should be carried out on the questionnaire collection process, including restricting each IP from submitting questionnaires repeatedly, and not allowing the submission of incomplete questionnaires. A total of 208 questionnaires were collected, of which 189 were valid, with an effective rate of 90.87%. In the sample, there is men accounting for 66.10%, 19-25 years old accounting for 76%, and junior college or above accounting for 98.30%. From the overall distribution of samples, the samples collected in questionnaire have certain rationality and can be used to represent the research object.

4. Data Analysis and Results

4.1. Common Method Deviation Test

Because the four variables of brand story, narrative transmission, social media interactivity, and consumer' purchase intention come from the same source, there may be common method bias in the data. Harman single factor test was used to test the severity of deviation of common methods. The results show that the variance of the first factor interpretation is 39.586%. The interpretation rate of the first factor variance is below 50%, and shows that the deviation problem of the common method is not serious.

4.2. Descriptive Statistics

Table 1 shows the mean value, standard deviation, and correlation coefficient of control variables and key variables. The results indicate that there is a significant positive correlation between brand story and consumers' purchase intention ($r = 0.717$, $p < 0.01$), and narrative transmission is significantly positive correlation with brand story ($r = 0.616$, $p < 0.01$) and consumers' purchase intention ($r = 0.607$, $p < 0.01$).

Table 1. Descriptive statistics and correlation analysis results of variables

	Gender	Age	EA	Occu-pation	DIM	PI	BS	NT	SMI
Gender	1								
Age	0.098	1							
EA	-0.168	-0.143	1						
Occu-pation	-0.089	0.209	-0.648**	1					
DIM	0.1	0.342**	-0.291*	0.431**	1				
PI	0.038	0.205	0.058	-0.076	0.118	1			
BS	-0.019	0.221	-0.012	0.04	-0.055	0.717**	1		
NT	0.042	0.116	-0.097	0.082	0.135	0.607**	0.616**	1	
SMI	-0.129	0.097	0.033	0.047	0.129	0.587**	0.579**	0.639**	1
MEAN	1.34	2.27	3.47	2.03	1.68	3.706	3.797	3.78	3.686
SD	0.477	0.582	0.537	1.485	0.88	0.678	0.836	0.773	0.765

Remark: ** At 0.01 level (double tail), the correlation is significant. * At 0.05 level (double tail), the correlation is significant.

4.3. Hypothesis Testing

The hierarchical regression analysis was used to test the intermediary hypothesis, and the results are shown in Table 2. After adding control variables, brand story has a significant positive impact on consumers' purchase intention ($M4, \beta = 0.884, p < 0.01$), so H1 has been verified by data. By testing the intermediary role of narrative transmission, brand story has a significant positive impact on narrative transmission ($M2, \beta = 0.723, p < 0.01$), and narrative transmission positively affects consumers' purchase intention ($M5, \beta = 0.712, p < 0.01$), and so H2 has been verified by data. To test the moderating effect of social media interaction, social media interaction can weaken the difference of the influence of brand story on consumers' purchase intention ($M9, \beta = -0.057, p < 0.01$).

To further verify the mediating effect of narrative transmission, the PROCESS plug-in was used to examine the significance of the mediating effect, and the number of replicate samples was set at 5,000. The analysis showed that the indirect effect of brand story on consumers' purchase intention through narrative transmission is 0.1670, with a 95% confidence interval of [0.0650, 0.3191], and the confidence interval did not include 0, that is, the negative emotion mediating effect is significantly present, and H2 is again validated.

The influence of brand story on consumers' purchase intention is interactively regulated by social media in H3. After controlling the related variables, it can be seen that there is a negative effect from Model 8 in Table 2, ($M9, \beta = -0.057, p < 0.05$) , which indicates that there is a moderating effect of social media interaction. In order to better explain the relationship between moderating effects, a simple slope analysis is used to draw a moderating effect diagram. As shown in Figure 2, social media interaction will weaken the influence of brand stories on consumers' purchase intention.

Table 2. Mediation effect test

	NT				PI				
	M1	M2	M3	M4	M5	M6	M7	M8	M9
Gender	0.064	0.021	0.087	0.034	0.055	0.029	0.087	0.034	0.053
Age	0.306	0.030	0.331	-0.007	0.110	-0.014	0.331	-0.007	-0.021
EA	0.112	-0.039	0.130	-0.054	0.088	-0.045	0.130	-0.054	-0.044
Occu-pation	-0.034	0.093	0.024	0.131	-0.016	0.110	-0.018	-0.141	-0.144
DIM	0.038	0.205	0.058	-0.076	0.118	-0.097	-0.024	0.131	0.127
BS		0.723**		0.884**		0.720**		0.884**	0.757**
NT					0.712**	0.227**			
BS*SMI									-0.057*
R2	0.098	0.539	0.121	0.781	0.579	0.805	0.121	0.781	0.793
ΔR2	0.098	0.441**	0.121	0.660**	0.558**	0.024*	0.121	0.660**	0.003*
F	1.150	10.130**	1.465	30.921**	14.587**	30.031**	1.465	30.921**	23.895**

Remark: ** At 0.01 level (double tail), the correlation is significant. * At 0.05 level (double tail), the correlation is significant. The coefficients in the table are standardized coefficients.

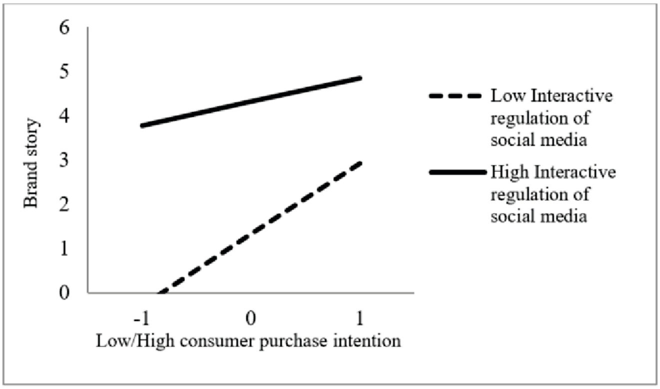


Figure 2. Interactive adjustment diagram of social media

5. Conclusion

This study explores the influence of brand story on consumers' purchase intention based on narrative transmission theory. The hypotheses are tested by empirical method. The results show that brand story can promote consumers' purchase intention, and at the same time, they are influenced by the intermediary role of narrative transmission and the interactive adjustment role of social media. Therefore, enterprises and marketers should pay attention to brand story, create excellent brand story, make full use of social media means, increase good brand story, and effectively promote consumers' purchase intention.

References

[1] Bauer H H, Grether M, Leach M. Building customer relations over the internet. Ind. Market. Manag. 2002 Feb, 31(2): 155-63.
[2] Liao M, Zhang J, Wang R, et al. Simulation research on online marketing strategies of branded agricultural products based on the difference in opinion leader attitudes. Info. Process. Agric. 2021 Dec; 8(4): 528-36.
[3] Ryu K, Lehto X Y, Gordon S E, et al. Effect of a brand story structure on narrative transportation and perceived brand image of luxury hotels. Tourism Manage. 2019 Apr; 71: 348-63.
[4] Granitz N, Forman H. Building self-brand connections: Exploring brand stories through a transmedia perspective. J. Brand Manag. 2015 Mar; 22(1): 38-59.
[5] Woodside A G, Sood S, Miller K E. When consumers and brands talk: Storytelling theory and research in psychology and marketing. Psychol. Market. 2008 Jan; 25(2): 97-145.
[6] Hong J, Yang J, Wooldridge B R, et al. Sharing consumers brand storytelling: Influence of consumers storytelling on brand attitude via emotions and cognitions. J. Prod. Brand Manag. 2021 Jun; 31(2): 265-78.
[7] Fahmy M, Ghoneim A. The impact of interactivity on advertising effectiveness of corporate websites: A mediated moderation model. Int. J. Market. Stud. 2016 Sep; 8(5): 41-55.
[8] Singh A. Effects of customer perception and social interactions on brand equity: Automobile sector. Int. J. Data Analy. Tech. Strate. 2017 Mar, 9(1): 75-98.
[9] Escalas J E. Narrative processing: Building consumer connections to brands. J. Consum. Psychol. 2004 Feb, 14(1-2): 168-80.
[10] Kang J A, Hong S, Hubbard G T. The role of storytelling in advertising: Consumer emotion, narrative engagement level, and word-of-mouth intention. J. Consum. Behav. 2020 Jan; 19(1): 47-56.
[11] Green M C, Brock T C. The role of transportation in the persuasiveness of public narratives. J. Pres. Soc. Psychol. 2000 Nov; 79(5): 701-21.
[12] Mullet G M, Karson M J. Analysis of purchase intent scales weighted by probability of actual purchase. J. Mark. Res. 1985 Feb; 22(1): 93-96.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230188

Research on the Construction of Evaluation Index System of Regional High-Quality Development Promoted by Intellectual Property Rights

Jiaming ZHONG[1], Lijuan YE and Xuan CHEN[2]
Xiangnan University, Chenzhou City, Hunan Province, 423000, China

Abstract. This paper expounds the connotation of high-quality development from the perspective of "five development concepts", and analyzes the dialectical relationship among innovative development, coordinated development, green development, open development and shared development. On this basis, a relatively complete evaluation index system of intellectual property to promote high-quality development has been constructed, and finally, the countermeasures of intellectual property to promote high-quality development have been proposed.

Key words. Intellectual property, five development concepts, high quality development, evaluation index system

1. Introduction

The report to the 19th National Congress of the Communist Party of China pointed out that the Chinese economy has shifted from a stage of high-speed growth to a stage of high-quality development, and is now in a crucial period of transforming the development model, optimizing the economic structure and shifting growth drivers. In this regard, the academic community has carried out a series of studies, forming a rich research results [1], such as the connotation of "high-quality development" [2], constraints and influencing factors [3][4], objectives and transformation path [5][6], index system construction and measurement [7]-[9] and other aspects of the research. Have carried out extensive research from theoretical or empirical perspectives [10]-[11].

In the 14th Five-Year Plan for National Economic and Social Development and the Outline of 2035 Goals, the work on intellectual property rights is centralized and separately formulated, and the strategy of making China stronger through intellectual property rights is clearly stated. In the main indicators of economic and social development in the 14th Five-Year Plan, the number of high-value invention patents

[1] The first two authors contributed equally to this paper and should be considered co-first authors;
[2] Corresponding author, Xuan CHEN, Xiangnan University, China; Email: wgyxycx@xnu.edu.cn
2021 Special Project of Intellectual Property Strategy Promotion in Hunan Province (Research on the Evaluation Index System of Intellectual Property Promoting High-quality Development in Hunan and Its Countermeasures) (Project No. : 2021F006K)

per 10,000 population is set. Judging from a series of national deployment around intellectual property work, the emphasis on intellectual property has been elevated to an unprecedented strategic height [1].

Practice shows that to promote high-quality development, on the one hand, we need to grasp the connotation of "high-quality development" scientifically; On the other hand, an appropriate evaluation index system of regional high quality development should be constructed to guide the practice of regional economic and social development.

Ma Ru et al. [12] constructed and measured the high-quality development evaluation index system of China's regional economy from five dimensions of high-quality supply, high-quality demand, development efficiency, economic operation and opening-up. Li Jinchang et al. [13], starting from the two aspects of "people's needs for a better life" and "unbalanced and inadequate development", the main contradictions of Chinese society in the new era, constructed a high-quality development evaluation index system consisting of five first-level indicators, that is ,economic vitality, innovation efficiency, green development, people's life, and social harmony, and 27 second-level indicators. Wei Min et al. [14] constructed a measurement system of high-quality economic development from 10 dimensions, including optimization of economic structure, innovation-driven development, efficient allocation of resources, improvement of market mechanism, stable economic growth, coordination and sharing among regions, quality of products and services, improvement of infrastructure, construction of ecological civilization, and benefits of economic achievements. Zhang Junyan et al. [15] pointed out that the essential connotation of high-quality development is to meet people's ever-growing needs for a better life, and to achieve efficient, fair, green and sustainable development, aiming at the existing problems of high-quality development index system and measurement in China.

There are few reports on the research of taking intellectual property elements as high quality development evaluation indicators.

2. Connotation and essence of regional high-quality development

Development is an important issue in today's society and the top priority for the Communist Party of China to govern and rejuvenate the country. An important task set forth at the 18th National Congress of the CPC is to "concentrate on construction, pursue development with undivided attention, grasp the law of development, innovate the concept of development, and solve difficulties in development". Centering on the concept of "innovative development", the Fifth Plenary Session of the 18th CPC Central Committee clearly put forward the five development concepts of "innovative, coordinated, green, open and shared". [11][16]

2.1 The Concept of "Five Major Developments"

(1) Innovative development is the driving force

Innovation is the soul of a nation's progress; it is the driving force for a country's prosperity and the source of eternal vitality for a political party. At present, from the perspective of international situation, the competition of comprehensive national strength is essentially the competition of innovation ability. From the perspective of domestic economic and social development, after more than 40 years of sustained and rapid development since the reform and opening up, China's economic aggregate has become the second largest in the world, with per capital GDP approaching 8,000 US dollars, but the problem of unbalanced and inadequate development is obvious. The Fifth Plenary Session of the 18th CPC Central Committee stressed that we must put innovation at the core of national development and continue to promote theoretical, institutional, scientific and technological, cultural and other innovations, so that innovation runs through all the work of the Party and the country and becomes a common trend in the whole society.

(2) Coordinated development is the method

Coordinated development is an important measure for scientific development, focusing on solving the problem of unbalanced development. Remarkable achievements have been made in China's coordinated development since reform and opening up. However, unbalanced, uncoordinated and unsustainable economic and social development still exist. The task of narrowing the development gap between urban and rural areas and between regions and promoting coordinated economic and social development remains arduous.

(3) Green development is our goal

Green development focuses on the harmonious coexistence between man and nature, and aims to build a beautiful China that is resource-conserving and environment-friendly. Green development determines the direction of development. We should properly strike a balance between economic development and environmental protection, firmly establish the idea that protecting the ecological environment means protecting productive forces, and improving the ecological environment means developing productive forces, and work more consciously to promote green, circular and low-carbon development.

(4) Opening up and development is a strategy

An important part of open development is the linkage between internal and external development. Since the beginning of the new century, economic globalization has deepened, economies are increasingly interdependent and interconnected, and mankind has become a community with a shared future. At the same time, the world economic structure is undergoing new changes, the international financial crisis has had a far-reaching impact, systemic and structural risks remain prominent, and China still faces grave challenges for its future development.

(5) Shared development is our destination

Social equity and justice are important issues that need to be solved. Fairness and justice are inherent requirements of socialism with Chinese characteristics, and achieving social fairness and justice is a consistent proposition of Chinese Communists and a major task in developing socialism with Chinese characteristics. The socialist cause is the cause of the vast majority of the people. The ultimate goal of development is to serve the people and ensure that everyone enjoys the opportunities and benefits of development. Therefore, sharing is an essential requirement of socialism with Chinese characteristics and an important measure to achieve fairness and justice.

2.2 The support of intellectual property rights to the concept of "Five Major Developments"

(1) Intellectual property is an important support for innovation and development. As a legal guarantee for innovation, intellectual property effectively stimulates innovation,

escorts innovation, and protects innovators from obtaining actual economic benefits. We should constantly improve the intellectual property system, promote the effective connection between intellectual property policies and policies of science and technology, industry, trade, finance and taxation, so as to promote building a powerful intellectual property country through institutional innovation [17][18].
(2) Intellectual property plays an important role in coordinating urban and rural development and regional development. In terms of balancing urban and rural development and narrowing the gap between urban and rural areas, the development of intellectual property rights has unparalleled room to play. Intellectual property rights should be used to promote agricultural modernization, narrow the development gap between urban and rural areas, reduce the Engel coefficient of rural areas, and coordinate urban and rural development.
(3) Green is realized in the course of development, which is the internal requirement of social progress. The key to green development is scientific and technological innovation. Environmental protection is increasingly dependent on the support of science and technology and the protection of intellectual property rights.
(4) Intellectual property trade is an important embodiment of intellectual property support for open development. Intellectual property trade promotes the international opening and sharing of innovative products, promotes open innovation, and improves the welfare level of the world.
(5) Increase the sharing of intellectual property between regions, as there is significant regional imbalance in the development of intellectual property. Increasing the transformation and application of intellectual property between regions will help optimize the allocation of intellectual property resources between regions, promote the circulation and sharing of intellectual property between regions, increase the social and economic benefits of intellectual property, as well as improve the level of social welfare.

3. The internal mechanism of intellectual property driving high-quality development

3.1 The innovative role of intellectual property rights

(1) Intellectual property provides the source power for innovation

① Intellectual property provides rules for innovation. The systematization and standardization of intellectual property data is the crystallization of human wisdom and the summary of objective laws, which provides a steady flow of power for innovation. To grasp and apply these wisdom and laws in practical activities, the innovation will form a virtuous cycle, so as to promote scientific and technological innovation, economic development and social progress[17][18].

② Intellectual property rights provide resource allocation for innovation. Intellectual property rights play an important role in providing big data support for innovation, saving costs for the early preparation and later protection of innovation achievements, establishing coordination mechanism for the profit distribution of innovation subjects, promoting the transformation of innovation achievements, comprehensively optimizing and integrating the allocation of production factors and innovation factors, and deeply integrating innovation achievements with economic and social development in various fields[17][18].

③ Intellectual property provides a platform for the dissemination of innovation results. Intellectual property is the bridge and link between innovation and market, and it is a very important link in realizing the value of innovation.

So it can be said that intellectual property provides a direct source of innovation.

(2) Intellectual property provides strong support for innovation

Among intellectual property rights, the invention patents with the highest technical content directly reflect the power of innovation ability. The quantity, quality, scale and level of intellectual property ownership, as well as the ability to use and manage intellectual property rights, have become important indicators to measure a country's economic, scientific and technological strength. Practice has shown that countries and regions with sound intellectual property systems and better implementation have stronger innovation capacity and higher level of economic and social development. For these countries, intellectual property is not only the choice of development strategy, but also the embodiment of comprehensive national strength.

(3) Intellectual property rights provide support for innovation

For enterprises, technological innovation is the key to win market advantages and improve competitiveness. Technological innovation cannot be separated from the protection of intellectual property rights and the incentive effect on innovative talents. If the protection of intellectual property rights is weak, technological innovation cannot play its role. Only by promoting scientific and technological innovation and unleashing and developing social productive forces can we achieve sustained and sound economic and social development.

3.2 Innovation-driven mechanism of high-quality development

(1) Technological innovation as the source of power to promote the continuous improvement of social efficiency.

In all kinds of innovation systems, technological innovation is always in the core position. From the micro point of view, technological innovation can upgrade the original production factors, improve the production efficiency, change the original production mode of enterprises, and reduce the dependence on resources. From a macro point of view, technological innovation has changed the original pattern of social resource allocation and made the overall allocation of social resources more efficient[19].

(2) Talent as the core driving factor to accelerate technology diffusion and application of achievements.

The core factor of innovation drive is talent drive, and human capital is the most active and positive factor in the whole process of innovation. Innovation needs a team, and the team needs innovative talents. Only with innovative talents can there be innovative achievements, and the promotion and application of innovative achievements must rely on talents. Therefore, from the macro level, the core of accelerating the construction of high-quality economic innovation drive is to strengthen the construction of innovative talents and consolidate the foundation of innovation human capital. From the micro level, innovative talents will produce innovative results, and the improvement of labor quality will improve labor productivity. At the same time, the progress at the micro level will promote the development at the macro level. With the innovation of talents, it will improve our independent innovation ability and strengthen our core competitiveness. Through the joint action of micro and macro level, talents become the core factor to promote high-quality economic development.

(3) Funding guarantees high-quality output.

Only high-quality innovation input can produce high-quality innovation, so as to improve the quality of innovation investment of Chinese government and enterprises, thus provide important guarantee for high-quality economic development.

(4) The synergy between the market and the government is related to the ultimate realization of high-quality development.

How to accurately and efficiently allocate innovation factors such as capital, technology and human resources to innovation entities is the key to achieve innovation-driven high-quality development.

3.3 Synergy between intellectual property rights and high-quality development

Synergetic theory holds that the system is composed of a large number of subsystems with completely different properties. Openness, interaction and self-organization form the spatial, temporal or functional structure of the system [20][21]. The key to the transition from disorder to order of the system lies in the interrelated "synergistic effect" [22][23] among subsystems within the system. Intellectual property rights and high-quality economic development can be seen as two closely related subsystems under the technology-economy system. Taking intellectual property as the variable that dominates the technology system and high-quality economic development [24], there is a coupling and coordination relationship between intellectual property and high-quality economic development, and the subsystem of intellectual property and the subsystem of high-quality economic development constantly exchange material, energy and information. Intellectual property subsystem provides continuous source technology supply for high quality development subsystem of economy; In turn, the subsystem of high-quality economic development provides sufficient R&D guarantee and market demand for the subsystem of intellectual property, and both of them promote and coordinate each other. When the new technology is applied to the market, or the demand of the high level of the market stimulates the emergence of the new technology, the technology-economy system will produce local mutations, promote the overall system transition from order to local disorder, and generate disturbances, and gradually make the overall system appear oscillating and even phased disorder until the sub-systems within the system complete self-reorganization and form endogenous variables. At the same time, under the joint action of other exogenous variables, the whole technology-economic system will return to an orderly state. The continuous evolution of the technology-economy system has been promoted by a new and more optimized orderly process from order to disorder within the technology-economy system [1].

4. Evaluation index system for the promotion of high-quality regional development by intellectual property rights

4.1 Construction principles of evaluation index system

A high-quality development evaluation index system should be established, and indicators should be screened according to the following principles [25].

(1) Guiding principle.

Based on the concept of high quality development, research and analysis of regional development prospects to guide the path of future regional development.

(2) Principle of comprehensiveness.

It can assess the level of regional economic and social development and existing problems in an all-round and multi-angle way.

(3) The principle of simplicity.

When selecting the evaluation index, it is required that the index be representative and can effectively reflect the development of the evaluation Angle.

(4) Principle of comparability.

In order to facilitate comparative evaluation, the selected indicators can be compared horizontally within the region.

(5) Operability principle.

In order to ensure the credibility and objectivity of the evaluation results, the data source is required to be authentic and reliable.

(6) Openness principle.

We should not only consider the commonalities of economic and social development, but also conform to the economic and social characteristics of different regions.

4.2 Evaluation index system of regional high-quality development promoted by intellectual property rights

On the one hand, for the high-quality development subsystem, the evaluation index system for regional high-quality development is studied from the five major development concepts. On the other hand, for the intellectual property subsystem, in the existing evaluation index system, many of them only consider the input or output of intellectual property as a single factor. However, from a performance perspective, the selection of evaluation indicators must cover the process and results of intellectual property activities. Therefore, this article studies the creation and application of intellectual property, and measures their indicators separately, thus form a relatively complete evaluation index system for promoting high-quality regional development by intellectual property rights. As shown in Table 1.

Table 1. Evaluation index system of regional high-quality development promoted by intellectual property rights

subsystem	Primary index	Secondary index	Index description
High quality development subsystem (Y)	Innovation-driven development (Y_1)	Innovation investment	R&D /GDP
		innovation output	Number of invention patent applications
	coordinated development (Y_2)	regional coordinated development	Per capita disposable income ratio of urban and rural residents
		Industrial Coordination Development	Proportion of added value of Tertiary sector of the economy in GDP
		risk-prevention	Rate of qualified Products
	Green development (Y_3)	Green Resources	Per capita water consumption
			Rate of harmless treatment of household waste
		Green Society	Energy consumption per ten thousand yuan of GDP
			Greening rate of built-up areas
		Green Management	Proportion of energy conservation and environmental protection expenditure
	Open development (Y_4)	International communication	Total number of people exit/entry
		international trade	Gross Import and Export Volume/GDP
			Total foreign investment/GDP

	Sharing development (Y_5)	Achievement sharing	Ratio of urban and rural consumption expenditure
			Ratio of per capita disposable income of urban and rural residents
			The ratio of regional per capita GDP to national per capita GDP
		public service	Per capita education expenditure
			Per capita healthcare expenditure
Intellectual Property Subsystem (X)	Intellectual Property Creation (X_1)	Investment of intellectual property	R&D personnel full-time equivalent
		Output of intellectual property	Number of invention patents per ten thousand person
	Application of Intellectual Property (X_2)	Transactions of intellectual property	Number of regional contracts for technology flow in the technology market
			Regional contract amount for technology flow in the technology market
		Maintenance of intellectual property rights	The proportion of effective invention patents to the total number of invention patents
			The rate of patent maintenance

$$\gamma(Y_i, X_j) = \frac{1}{n}\sum_{1}^{n} \frac{\min\min\left|y_i(k) - x_j(k)\right| + \zeta \max\max\left|y_i(k) - x_j(k)\right|}{\left|y_i(k) - x_j(k)\right| + \zeta \max\max\left|y_i(k) - x_j(k)\right|} \tag{E1}$$

[26]

According to Table 1 and (E1), we can obtain the grey correlation $\gamma(Y_i, X_j)$ between the high-quality development subsystem indicator Yi (i=1,2,3,4,5) and the intellectual property subsystem indicators X1 and X2, thereby obtaining the degree of influence of the intellectual property subsystem indicator Xj (j=1,2) on the high-quality development subsystem indicator Yi (i=1,2,3,4,5).

5. Measures for intellectual property to promote high-quality development

5.1 Accelerate core technology innovation and rationally allocate science and technology resources.

First of all, we should strengthen the investment in basic research, increase the proportion of basic research funds, establish a reasonable investment mechanism for technological innovation, and rationally plan the allocation of various scientific and technological elements. Secondly, we should establish and improve our intellectual property protection system and stimulate the innovation vitality of enterprises, universities and research institutes. Finally, we should deepen the supply-side reform, deeply study the current input and output structure of science and technology in our country, promote the application and popularization of scientific and technological innovation, and apply the core technologies to push our industrial chain towards high-end development[19].

5.2 Strengthen the training of innovative talents and stimulate the vitality of innovation.

Establish and improve the talent management mechanism to meet the requirements of innovation and development and conform to the law of innovation. Deepen the reform of education system, do a good job in the top-level design of talent training programs, innovate education methods, improve the talent evaluation system, rationalize the evaluation indicators of innovation ability, quality and contribution, stimulate the enthusiasm of innovative talents [27], accelerate the training of strategic innovative talents with international frontier level, and provide the core power for the high-quality development of Chinese economy.

5.3 Increase financial support for innovation to provide strong guarantee for innovation activities.

The government's investment of funds should improve the efficiency of the use of funds to avoid the waste of funds; Enterprises should establish and improve the management methods of capital investment, enhance their macro-planning ability and market competitiveness, and avoid scattered allocation of scientific and technological resources and low overall operation efficiency caused by low efficiency of capital investment within enterprises.

5.4 The government and the market should work together to promote development.

The market should play a decisive role in the allocation of innovation resources, while the government should also provide good management and quality services. Finally, through the synergy between the market and the government, as well as the effective supply and allocation of innovation factors and resources, we can realize the high-quality development of innovation-driven economy[28].

Acknowledgements

It is the research result of The Key Cultivation Base for "The 14th Five-Year Plan" of Educational and Scientific Research (Lifelong Education Research Base(Fundamental Theory Area)) in Hunan Province (XJK22ZDJD58).

Reference

[1] Liu FC, Lin Y. Coupling coordination between in tellectual property and regional high quality development: anempirical analysis of provincial administrative regions in China[J]. Journal of Central China normal University(Nat. Sci), 2021, Vol. 55, (05): 717-726.
[2] Jin B. Study on the “high-quality development” economics[J] China Industrial Economics, 2018,(4): 5-18.
[3] Wang HY, Li XY, Xu YL. Research on performance evaluation ad influencing factors of high-quality economic development driven by scientient ificand technological innovation in China[J]. Economist, 2019, (11): 64-74.
[4] Ren B P, Wen F A. The criteria, determinant sand ways to achieve high quality development in China in the newera[J]. Reform, 2018, (4): 5-16.
[5] Zhang JK, Hou YZ, Liu PL, etal. The goal and strategic path of high quality development[J]. Management World, 2019, 25(7): 1-7.

[6] An SX. A Research on the path of promoting high-quality economic development: a literature review[J]. Contemporary Economy&Management, 2018, 40(9): 11-17.
[7] Ma R, Luo H, Wang HW, etal. Study of evaluating high-quality economic development in Chinese regions[J]. China Soft Science, 2019, (7): 60-67.
[8] Li JC, Shi LM, Xu AT. Probein to the assessment in dicator system on high-quality development[J]. Statistical Research, 2019, 36(1):4-14.
[9] Wei M, Li SH. Study on the measurement of economic high-quality development level in China in thenewera[J]. The Journal of Quantitative&Technical Economics, 2018, 35(11): 3-20.
[10] Wang YQ, Guo YB, Zhao Y, etal. Coordinated Regional developement driven by scientificand technological innovation: theoretical basis and Chinese practice[J]. China Soft Science, 2017,(11): 86-100.
[11] Ou JF, Xu CJ, Liu YQ. The Measurement of High-Quality Development Level from Five Development Concepts：Empirical Analysis of 21 Prefecture-Level Cities in Guangdong Province[J].
[12] Ma R, Luo H, Wang HW, etal. Study of evaluating high-quality economic development in Chinese refions[J]. China Soft Science, 2019, (07): 60-67.
[13] Li JC, Shi LM, Xu AT. Probe in to the assessment indicator system on high-quality development[J]. Statistical Research, 2019,36(01): 4-14.
[14] Wei M, Li SH. Study on the measurement to feconomic high-quality development level in China in thenewera[J]. The Journal of Quantitative&Technical Economics. 2018, 35(11): 3-20.
[15] Zhang JK, Hou YZ, Liu P L, etal. The goal and strategic path of high quality development[J]. Management World, 2019, 35(07): 1-7.
[16] Qin X, The Dialectical Relationship between Five Development Concepts [C]. Guangming Daily, February 04, 2016.
[17] Hu CP. Research on the Mechanism of Patent Intensive Industry Promoting High-Quality Economic Development[J]. Henan Science and Technology, 2022,41(10): 133-136.
[18] Gu XY. Intellectual Property Measures to Promote People's Livelihood Happiness: A Perspective of Five Development Concepts[J]. On Economic Problems, 2016(12): 7-10+16.
[19] Zhang ZH, Guo X, Yi L. Innovation-Driven Mechanism of High-Quality Economic Development[J]. Journal of Xi'an Jiaotong University(Social Sciences), 2019,39(06): 39-46.
[20] Haken H, Wunderlin A, Yigitbsi S. On the foundations of synergetics[C]//WEINGARTNER P, SCHURZ G. Law and prediction in the light of Chaos research. Lecture Notes in Physics. Berlin: Springer, 1996: 473.
[21] Haken H. Visions of synergetics[J]. International Journal of Bifurcation and Chaos, 1997,7(09): 759-792.
[22] Wu YM, Lang DF, Zhang ZH, etal. Coordination degree model of environment economy system and its application[J]. China Population, Resources and Environment, 1996,(02): 51-54.
[23] Meng QS, Han WX. Study of the coordination measurement model with respect to composite system[J]. Journal of Tianjin University(Science and Technology), 2000, 33(4): 444-446.
[24] Arthur WB. Technologies, increasing returns, and lock in by historical events[J]. Economic Journal, 1989, 99: 116-131.
[25] Cheng QY. Construction and Application of High-quality Development Evaluation Index System[J]. Statistics & Decision, 2022, (24): 28-32.
[26] Huang GH, Fang G. System Engineering Methods and Applications[M]. Guangzhou: Jinan University Press, 2006.
[27] Gao SG. Research on the Core driving factors of achieving high-quality development [J]. Macroeconomic Management, 2018, (9): 63-68+77.
[28] Liu SJ. Implementing the decisive role of market in resource allocation [J]. Economic Research, 2014, (1): 11-14.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230189

A Survey on the Status Quo of Sex Education for Left-Behind Children in Rural Areas: A Case Study of Rural Areas in Southern Hunan Province

Lanyan Qin, Jie Ouyang, Xuli Li, Linzi Qin[1]
Xiangnan University, Chenzhou, Hunan, 423000, China

Abstract. Objective: To understand the status quo of sex education in left-behind children in rural areas. Method: Self-compiled questionnaires on sex education for left-behind children, teachers and general public and an interview questionnaire were administered to the target respondents. A total of 471 valid copies of questionnaire were retrieved, along with some interview data. Results: (1) There was a general lack of sex knowledge among left-behind children and hence a deficiency of the ability and consciousness to protect themselves; (2) School-based sex education was inadequate, and teachers believed that sex education should be strengthened at school; (3) There was a serious lack of family sex education, while the general public showed growing concern for sex education. Conclusion: It is necessary to strength sex education among left-behind children in rural areas from various dimensions, so as to elevate their sex knowledge and awareness about sexual health. The coordinated promotion by schools, families and society is an inevitable choice and an important measure to improve the effect of sex education.

Keywords. rural areas, left-behind children, sex education

1. Introduction

Along with the acceleration of China's urbanization, a large number of rural residents move to urban areas, resulting in an increasing number of left-behind children in rural areas [1]. Left-behind children are children who remain in rural regions of the country while both parents leave home to earn a living or one parent leaves home and the other cannot exert guardianship. These children are juveniles aged below 16 with rural residency and do not live with their parents [2]. Compared with their urban counterparts, left-behind children in rural areas are much more vulnerable to sexual assault due to the lack of guardianship, family care and protection, and children's inability to protect themselves and discern sexual assault[3]. The urgent task is to conduct surveys on the status quo of sex education for left-behind children in rural areas and help them establish a self-defense mechanism against sexual assault. We conducted a series of surveys and researches in rural areas of southern Hunan Province,

[1] Corresponding Author, Linzi Qin, Xiangnan University, Chenzhou City, Hunan Province, 423000, China; Email: 3158342060@qq.com.

aiming to understand the status quo of sex education and sexual cognition among left-behind children in rural areas and analyzing the problems existing in sex education among left-behind children. We intended to find new solutions to the socialization of sex education for left-behind children. A sex education platform for children may be built in the future based on our survey findings. The current survey notifies the general public about the importance of developing effective approaches to sex education for children and finally contributes to the healthy growth of children in rural areas, thereby promoting the development of rural education and rural culture construction.

Researches on the problem of left-behind children is mainly focused on basic education, mental health, individual safety, types of guardianship, emotional life, socialization difficulties, scale, distribution and left-behind girls, but there is little research literature on "sex education for rural left-behind children". It can be seen that the current domestic research on this area needs to be improved, and specific feasible countermeasures need to be proposed for the problems related to sex education of rural left-behind children.

Sex education is an important course concerns the health and happiness of children throughout their lives. After a century of exploration and development, systematic theories of sex education has formed in some foreign developed countries. A comprehensive education system of sex education content has established and was integrated into the teaching of different age groups[4].Cheng Zizhen (2017) argues that rural children have a great demand for sex education during puberty, but the main bodies such as families, schools and village committees are in the absence of sex education for children, guardians lack awareness of sex education, and school sex education is vacant[5].Liao Qinxiang (2019) argues that left-behind children lack systematic and reasonable ways to actively acquire sex knowledge, and lack the necessary self-protection awareness when they encounter sexual abuse[6]. LiYi (2022) believes that the primary strategy to prevent sexual assault is to raise students' awareness of sexual assault prevention. In terms of intervention mechanism, a multidimensional system should be constructed to identify potential offenders for early detection and prevention. Actively promote sexual assault prevention methods and the laws[7].Zhang Xiaobing (2019) believes that sexual assault prevention education for parents and protection for left-behind children should be strengthened, emphasizing the refinement of responsibilities in terms of prevention and control countermeasures[8]. Liu Renfeng (2019) believes that education on children's sexual knowledge should be strengthened while strengthening the punishment for sexual crimes against left-behind children[9].

These studies have contributed to the development of children's sex education, but there are also some shortcomings. The research on rural children's sex education must gradually deepen to the theoretical level, and carry out research at the practical level to reveal the essence of the problem of sex education for left-behind children, find out the crux and prescribe the right remedy.

2. Survey design: respondents, methodology and contents

2.1 Respondents

Our survey was conducted among left-behind children in rural areas of southern Hunan Province. Located in central China and bordering with Guangdong, southern Hunan is geographically dominated by mountainous areas and backward in education and economy. The coastal areas of Guangdong Province, which are economically developed and geographically adjacent to southern Hunan, are the destination of many migrant workers from southern Hunan Province. For this reason, left-behind phenomenon is widespread in southern Hunan, and left-behind children in rural areas of southern Hunan constitute representative samples for our survey over this topic.

2.2 Methodology

Our survey was a combination of questionnaire survey and face-to-face interviews. Different questionnaires were compiled for teachers, general public, children, and parents. The questionnaire for students was designed to gain a preliminary knowledge about children's growth and development, sexual consciousness, status quo of sex education, and sexual attitude. The questionnaire for teachers was to assess teachers' mastery of sex knowledge, their attitudes toward sex education, and the status quo of sex education at school. Face-to-face interviews were conducted among different adult groups regarding their understanding of sex knowledge, their attitudes towards sex education, and ways to administer sex education. SPSSAU was used for questionnaire data analyses, which included descriptive statistics, comparison of means, t-test, and correlation analysis. The survey findings can be summarized as follows:

2.3 Contents of survey

This survey was conducted in light of the reality of sexual safety among China's left-behind children in rural areas at the present stage and the undergoing policy advocacy for protection against sexual assault. We first expound on such core concepts as sex education and sexual safety education and discuss the status quo and influence factors of sex education for children. Our findings provide theoretical support for children's sex education in China. With basic knowledge acquired by questionnaire survey and interviews regarding the status quo and needs of left-behind children in rural areas, we plan to initiate campaigns to spread sexuality knowledge among left-behind children through group activities. Our experience will be valuable for delivering sex education among children in rural areas.

3. Survey results and statistical analysis

3.1 Status quo of sex education for left-behind children in rural areas

a. Sexual physiology of left-behind children in rural areas

Sexual development is considered a hallmark of adolescence and implicates a variety of sex-related factors, including sexual psychology, sexual consciousness, and sexual behaviors. In this study, the status of sexual development in girls was assessed based on secondary sex characteristics, including menarche and other physical changes. According to our survey, girls whose menarche occurred at an age of 11 to 12 years old accounted for 9.93% and 23.57%, respectively. Those who begin sexual development in this age accounted for no small proportion.

But due to an absence of sex education, most girls felt the panic upon menarche. The respondents who turned to their mothers for help accounted for the largest proportion. According to many respondents, their mothers were not around when menarche occurred because they were left-behind children. This is a typical example of inadequate family sex education.

b. Mastery of sex knowledge of left-behind children in rural areas

Our survey showed that most left-behind children did not have a correct understanding of their private parts. Neither did they know anything about sexual behaviors. Without a clear knowledge about basic concepts about sex, these children can hardly protect themselves in the face of sexual assault if any of these things happen, because they are unaware that something bad has happened to them. There was also a shortage of knowledge about pregnancy and contraception. When asked about the reasons for women getting pregnant, only a few girls gave a correct answer. Most girls said that they did not know anything about "the impact of induced abortion on girls". As for the knowledge about contraception, barely no one knew anything more than the very basics.

c. Analysis of the status quo of sexual safety in left-behind children in rural areas

Nothing is more important than to instill the sense of sexual self-defense in children as far as sex education is concerned. Once being assaulted sexually, children are very likely to suffer serious consequences, such as anger, anxiety, depression, fear and personality disorders. One question in the survey was concerned with the status quo of the consciousness of sexual self-defense. The question went like this: Can any other person, including your parents, touch your breasts or private part without your permission? Some children chose "They cannot do so in any circumstances." Other girls said that their parents can, and there were also girls who said they were unclear about it. We can see that few children know how to react in the face of sexual assault. Another question was "Do you know how to protect yourself if sexual assault occurs?" More than 40% of the children surveyed were not sure whether they could protect themselves or not.

As for knowledge about sexually transmitted diseases and AIDS, most children said that they knew nothing about them or they only heard somebody mentioned it or felt these concepts quite obscure. As for how AIDS is transmitted, most children said they did not know. It is easy to see that if children do not know enough about sexual self-defense before they step into society, they are very likely to face sexual health hazards. Our survey has fully demonstrated the urgency and necessity of administering sex education for children.

d. General lack of common knowledge in law among left-behind children in rural areas

Our survey has indicated a widespread lack of common knowledge in law among the respondents. One question went like this: "Do you know that having sex with young girls aged below (?) constitutes a violation of law, whatever the cues used or regardless of whether the girls themselves consent to it or not?" The respondents gave various answers as to the minimum age of young girls considered legally competent to consent to sexual acts. Very few respondents chose the right answer, and some of them might have just "guessed right". Apparently, children in rural areas had a very low level of legal knowledge regarding sex.

3.2 Teachers' cognition about sex education for children

Most teachers surveyed (73.33%) mastered more sex knowledge than students due to their senior age and rich social experience. Only a few teachers said that they only knew a little about sex (24.44%) or nothing at all (2.22%). Teachers' mastery of sex knowledge also affected teachers' attitudes to sex education at school. We observed that 92.22% of the teachers surveyed believed that it is necessary to offer sex education at school, which is beneficial to the physical and psychological growth of students; 5.56% of the teachers believed that it is only of moderate importance to offer sex education at school; 2.22% of the teachers believed that sex education is not important at all. Besides, teachers' mastery of sex knowledge was related to the practice of teaching and discussion about sex-related topics with their students.

Table 1. Pearson's correlation-detailed format

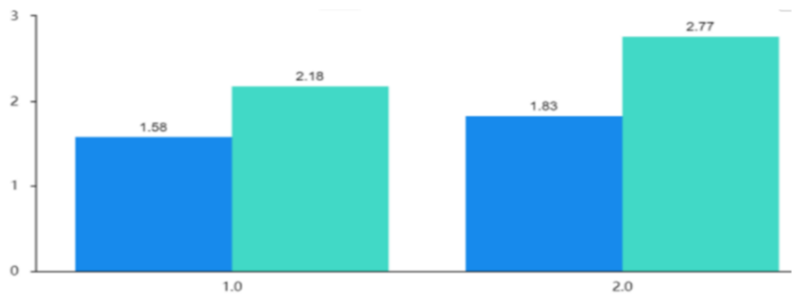
		4. Have you ever learned anything about sex?
9. Have you ever communicated individually with left-behind students in your class about sex?	Correlation coefficient	0.161
	p-value	0.13
7. Do you spare some time in your class to teach about sex? If yes, how many classes are held per week for sex education?	Correlation coefficient	0.219
	p-value	0.038
12. What is your attitude when girls consult with you about sex?	Correlation coefficient	0.279
	p-value	0.008

The intensity of correlation was measured using Pearson's correlation coefficient. Whether the teacher has known enough about sex and whether the teacher has communicated individually with girls about sex.

The intensity of correlation was measured using Pearson's correlation coefficient. The coefficient of correlation between whether the teacher has known enough about sex and whether the teacher has communicated individually with girls about sex was

0.161, which was close to 0. Besides, the p-value was 0.130, which was higher than 0.05, indicating no significant correlation between the two. The coefficient of correlation between whether the teacher knows enough about sex and the frequency of sex education classes per week was 0.219, and the significance level was set to 0.05. Therefore, the two variables had a significant positive correlation. The coefficient of correlation between whether the teacher knows enough about sex and the teacher's attitudes towards students' consultation about sex was 0.279, and the significance level was set to 0.01. Therefore, the two variables had a significant positive correlation.

As seen from above, teachers with a greater reserve knowledge about sex are more capable of helping students learn about sex and encouraging them to know more about their bodily development and the changes occurring to their appearance. More communication with students about sex can help prevent inferiority complex due to bodily development and raise their awareness of sexual assault and self-defense.



9. Have you ever communicated individually with left-behind students in your class about sex?

10. Do girls in your class read books about sex-related topics in their spare time?

Figure 1 Teachers help students learn about sex

Generally, the number of left-behind children in each class varies. According to our survey, 61.11% of the teachers surveyed paid home visits to left-behind children to know about their family background and physical and mental conditions. Nevertheless, teachers still do not pay enough attention to sex education for left-behind children in rural areas. Only 32.22% of the teachers surveyed said that they had communicated individually with left-behind children about sex; 58.89% of the teachers did not know whether their students read books about sex-related topics in their spare time or not. But most teachers (65.56%) agreed that there was a general lack of sex knowledge among their students; 57.78% of the teachers surveyed would communicate with their students if they were consulted about sex and answer their doubts; 34.44% of the teachers would only provide simple answers to questions raised by students about sex; only 7.78% of the teachers said that they refused to answer any questions about sex. Teachers' attitude towards sex also had an impact on students. If teachers refuse to talk about sex, the students may feel that sex is a bad thing. In that case, bodily development during adolescence may undesirably increase the likelihood of inferiority complex. When assaulted sexually, the children tend to keep their mouth shut or get threatened and suffer from long-term sexual assault. Teachers should provide positive guidance for students when being consulted about sex and make efforts to ensure their physical and mental health.

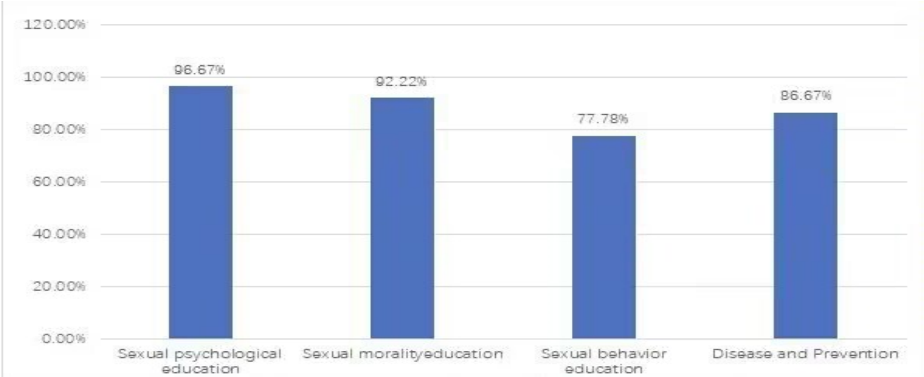


Figure 2. Teachers believed that sex education should be strengthened at school

Most teachers believed that the current sex education is far from comprehensive at school. According to their opinions, schools should strengthen education in sexual psychology, sexual morality, sexual behaviors, and sexually transmitted diseases and their prevention. Teaching activities of various kinds can be conducted at school. Teachers should also improve their sexual literacy to better help left-behind children in rural areas learn about sex.

3.3 A survey on public attitudes towards sex education for children

Table 2 Results of ANOVA

	8. Do you think that it is necessary for children to learn more about sex? (mean ± standard deviation)			F	p
	Necessary (n=3)	Unnecessary (n=297)	Let it be (n=81)		
7. Have you ever paid any attention to sex education for children?	1.67±0.58	2.19±0.50	1.89±0.45	13.103	0

Analysis of variance is to compare differences of means in quantitative variable Y with respect to categorical variable X. As seen from the table above, one-way ANOVA was used to compare the difference in the answers given to the above two questions among adults. The public attitudes towards the necessity of sex education for children and whether the adults have paid any attention to sex education for children were significantly correlated with each other (P<0.05). There seemed to be a consensus among the general public on the widespread absence of parents in sex education and the urgency of promoting sex education for children. Further analysis showed that the necessity of family sex education was significant at the 0.01 level (F=13.103, p=0.000). The mean score was compared between the groups, and the score for “Necessary” was higher than that for “Let it be” (see the figure below).

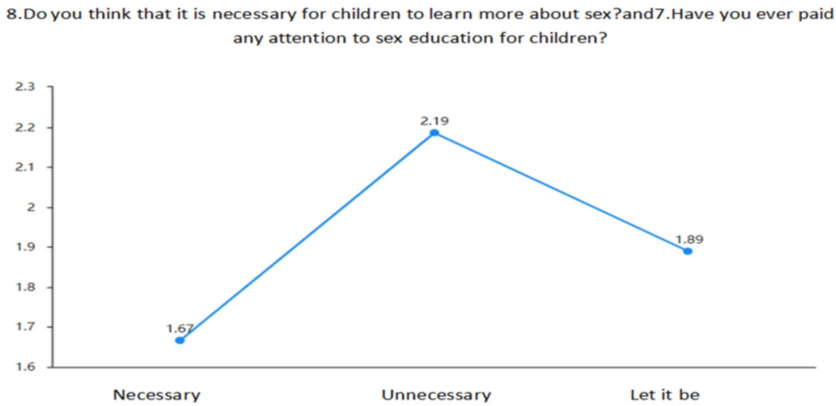


Figure 3 The public concern for sex education

Most adult respondents surveyed believed that children should learn more about sex. Very few respondents said that they did not consider it necessary. The above findings indicate that the public shows a growing concern for sex education for children.

Table 3. Results of t-test

	4. Have you ever learned anything about sex? (mean ± standard deviation)		t	p
	Yes (n=348)	No (n=33)		
5. Have your parents ever told you anything about sex? (please answer this question based on your personal identity)	2.35±0.50	2.06±0.24	5.817	0

As shown from the table above, t-test (independent-samples t-test) was performed to determine whether children's mastery of sex knowledge was significantly correlated to parents administering sex education to their children ($P<0.05$). We observed a significant correlation at the 0.01 level ($t=5.817$, $p=0.000$). It was also found that the average score of parents administering sex education to their children (2.35) was much higher than that of parents not administering sex education to their children (2.06).

The assumption made for analyzing the differences in answers to two questions, "Do you know anything about sex?" and "Have your parents taught you anything about sex?" using the t-test (independent samples t-test) was that children know about sex primarily from their parents. A P-value below 0.05 indicated that the two were significantly different. Long absence of parents means that the parents cannot teach their children about sex, which reduces the possibility of their children learning anything about sex. In China, the reality is that the long-term absence of parents is a widespread problem in rural areas, leading to a lack of sex education among left-behind children. This in turn has an adverse impact on children's psychological and physical health. Therefore, it is necessary to discuss how to promote sex education in left-behind children in rural areas.

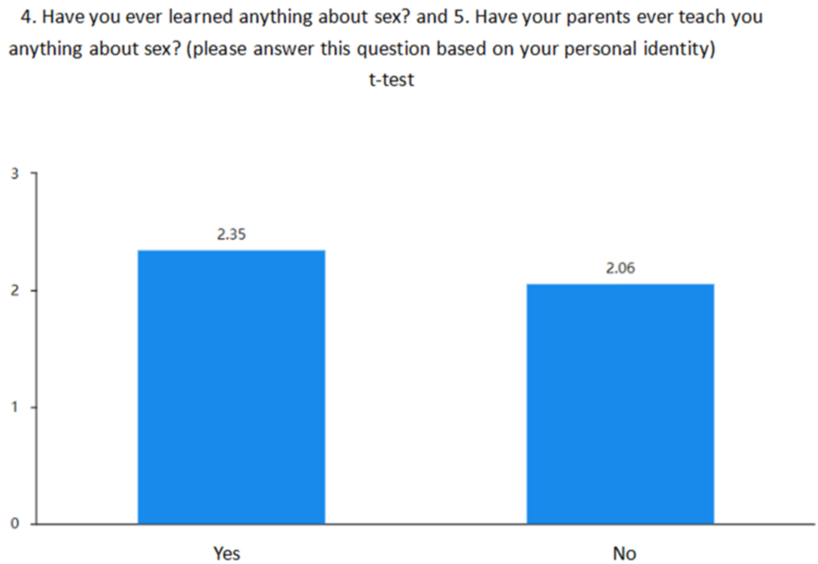


Figure 4 Children's learn about sexuality from their parents

4. Discussion

We arrive at the following conclusions based on our survey:

(1) Left-behind children in rural areas generally have limited sex knowledge. Without their parents by their sides, these children learn sex knowledge primarily from their peers or the Internet. We observe a lack of knowledge about sexual assault and how to prevent it among these children. They barely know anything about laws related to sexuality, which agrees with previous findings.

(2) Few schools provide sex education through specific curriculum. Teachers are more positive about sex education than before and consider it necessary for schools to provide sex education classes and launch other activities to teach sex knowledge.

(3) There is a serious lack of family sex education. Left-behind children do not live with their parents for an extended period of time, and they usually feel alienated from their parents. These children have little chance to learn about sexual health and care and how to protect themselves against sexual assault. Therefore, they are more vulnerable to physical and psychological hazards in relation to sex. The general public shows growing concern for sex education, which is a sign of social progress.

5. Conclusion and Recommendation

We have identified several problems existing in sex education among left-behind children in rural areas through our survey. Such as children’s lack of sex knowledge, school and family sex education is in the state of absence [5, 6].This is consistent with previous research , indicating that sex education for rural children 's has not made much progress. Scholars suggested that strengthen the sex education and the positive publicity role of the media should be brought into play[9,10]. However, we found that most of teachers believed that sex education should be strengthened at school and hope to take various forms to carry out it .And the public 's attitude towards sex education is becoming more and more open and positive. This shows the development of society and the progress of thought. These changes will bring positive help to the development of sex education.

Children grow up healthily only in a healthy environment [11]. Education is interrelated and interpenetrated. We should not only start from individual problems, but also find out the problems of each system from the perspective of ecosystem and the social structure in which they are located[12].

It is necessary to pay attention to the integration and mobilization of various social resources, form a linkage mechanism of mutual coordination[13].The coordinated promotion by schools, families and society is an inevitable choice and an important measure to improve the effect of sex education. We should step up our efforts to strengthen sex education among left-behind children in rural areas. which include the following aspects:

a. The family should pay more attention to children's physical and psychological health. In the first place, parents should have correct sex knowledge and a keen awareness of sexual health themselves. They should attach due importance to children's physical and psychological health and provide sex education for their children.

b. The schools should provide sex education classes. They should assign specific teaching staff for sex education and provide sex education classes to spread sex knowledge and cultivate awareness of sexual health. It is necessary for schools to enrich and diversify the teaching materials and teaching method as far as sex education is concerned.

c. The government should accelerate the construction of policies and legal system regarding sexuality. Specifically, the government should release more policies and laws to fight sexual assault against left-behind children and supervise the implementation of relevant laws to improve the defense mechanism and safeguard the physical and psychological health of left-behind children.

d. Concerted efforts should be made to establish a sex education model involving multiple parties. Schools, families, government, and communities should work together to build a multi-dimensional system that promotes sex education and elevate awareness of self-protection against sexual assault among left-behind children in rural areas.

Acknowledgements

Thanks for the funds from2021Xiangnan University project: ([2021] No. 139, 2021XJ51). 2022 Chenzhou social science project (NoCZSSKL202205). 2022 Hunan Social Science project: Research on the Current Situation of Sexual Safety Education for Rural Left behind Girls and the Prevention Mechanism of sexual assault (22YBA234), 2022 the Innovation and Entrepreneurship Training Program for College Students in Hunan Province: Research and Practice on the Integration of Red Culture into Primary and Secondary School Education and Teaching under the Background of "Internet plus" (XJT [2022] No. 174: 4302).

References

[1] The Ministry of Civil Affairs released data showing that rural left-behind children are more than 20 % less.[EB / OL][201811-02].https://www.gov.cn/guowuyuan/2018-11-02/content_5336714.htm
[2] Opinions of the State Council on Strengthening the Care and Protection of Rural Left-behind children in China[EB / OL][2016-02-29].https://www.gov.cn/gongbao/content/2016/content_5045947.htm
[3] Huang Baozhen, Zheng Xinying, Jin Yang, Yang Qi. Investigation and analysis of the current situation of sexual education and sexual abuse cognition of left-behind children in cities [J].China Sex Science, 2020, 29 (08) : 149-153.
[4] Ye Songqing. Research on the education and care mechanism of rural left-behind children in Anhui Province-Taking Feixi County of Hefei City as an example [J].Journal of Huaibei Normal University (Philosophy and Social Sciences Edition).2017 (05) : 51-58.
[5] Cheng Zizhen. Adolescent sex education for left-behind children in rural areas - A case study of northwest Guangdong [J]. Southern Journal, 2017 (01) : 83-85.
[6] Liao Qinxiang. Research on the current situation and countermeasures of sex education for left-behind children in rural China [J].Rural economy and technology, 2019,30 (13) : 208-210.
[7] Li Yi. Prevention strategies and intervention mechanisms for sexual abuse of left-behind children in rural areas [J].International Public Relations, 2022 (10): 71-73.DOI : 10.16645 / j.cnki.cn11-5281 / c.2022.10.009.
[8] Zhang Xiaobing. Characteristics, difficulties and solutions of sexual assault cases of left-behind children in rural areas [J].Application of law (judicial cases), 2019 (04) : 96-107.
[9] Liu Renfeng, Peng Jing. Research on sexual safety and education of left-behind children in rural areas-Based on the survey of some rural primary and secondary schools in Ningxiang City, Hunan Province [J].Education Observation, 2019, 8 (08): 54-56 + 60.DOI : 10.16070 / j.cnki.cn45-1388 / g4s.2019.08.02.
[10] Cao Bo. Empirical Study on Risk Prevention and Control of Sexual Assault on Minors in Southwest China - - Based on the Analysis of 93 Cases of Hot Media Reports [J].Guizhou Social Sciences, 2020 (12) : 89-96.DOI : 10.13713 / j.cnki.cssci.2020.12.013.
[11] Li Yifeng. Practical research on social work to improve the learning motivation of left-behind children in H village from the perspective of ecosystem theory [D].Guizhou University, 2022.DOI : 10.27047 / d.cnki.ggudu.2022.002633.
[12] Qiu Chenglin. Study on case social work of rural children in distress from the perspective of ecosystem theory [D].Northwest University of Agriculture and Forestry, 2020.DOI : 10.27409 / d.cnki.gxbnu.2020.000664.
[13] Xiao Lei, Chen Qiaoling, Duan Guimin. Research on the current situation and countermeasures of sexual health education in rural primary schools in Sichuan Province [J].China Sexology, 2019, 28 (07) : 157-160.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230190

Performance Evaluation Framework for Emerging Statistical Index Based on Mutual Spectral Analysis

Meijiao WANG[1], Yanjun QIAN[2,a] , Libo HE[1], Yunyun WU[1]

[1]*Yunnan Police College, Kunming, China*
[2]*Survey office of the National Bureau of Statistics in Yunnan,China*

Abstract. With the continuous development of emerging economic forms, the disadvantages of traditional statistical measures are becoming increasingly prominent. The performance measure of the emerging economic statistical index is also an urgent problem to be solved. This paper propose a performance evaluation frame for emerging economic statistic index which takes traditional index as the benchmark. The proposed method uses the spectral analysis method to analyse the periodogram of the time series, determine the appropriate parameters according to the filter, and filter the time series to obtain the trend and periodic terms of the sequence. Use the ANOVA method for trend items to analyse the consistency of trend items, and use the mutual spectral analysis method to analyse the consistency and leading lag of new indicators and benchmark indicators. By using the method of frequency domain analysis and mutual spectral analysis, the components of this frame include the preprocessing, analysis, comparison of innovative statistical index data, and obtain the performance evaluation results of innovation indicators, which provides a solution for the quantitative analysis of innovation index. This paper also give a specific example to present how this framework works to evaluate the performance of an emerging economic statistic index.

Keywords. ANOVA, Statistical Index, Frequency Domain Analysis

1. Introduction

Nowadays, the world is undergoing a new round of scientific and technological revolution and industrial transformation. Innovation-driven, fostering and expansion of

[a] Corresponding Author: Yanjun QIAN, Survey office of the National Bureau of Statistics in Yunnan, No.95 Huancheng East Road, Panlong District,Kunming, China; E-mail:q_y_jun@163.com.

new industries, new forms of business, and new models are the space and potential for economic development [1]. To quantitatively measure the economic operation status, it is necessary to systematically sort out the definition, characteristics and standards of various economic indicators under the new situation, and to explore and innovate the index system, evaluation system and statistical methods [2]. At present, quantitatively evaluating the performance of statistical indicators obtained by innovation becomes an urgent problem to be solved.

Traditional analysis methods include ANOVA, coefficient of variation analysis method, econometric model, etc. The coefficient of variation is the ratio of standard deviation to mean of the raw data. The coefficient of variation analysis method can be used when the two sets of data discrete magnitude need to be compared, and the data measurements vary widely, or the data scale is different. This method is a common indicator to measure the degree of regional development imbalance [3]. For some certain economic fields, the coefficient of variation can also be used to comprehensively evaluate the index system [4].However, the coefficient of variation is sensitive to perturbation and lacks accuracy which may require the manual intervention processing.

ANOVA was able to identify the magnitude of controllable influence on the findings by calculating the contribution magnitude of variants from different sources to the total variation. For example, Huang Xiu et al. studied the degree of difference between ASPI and CPI of Alibaba online purchase price index by ANOVA [5]; investigated the influencing factors of the domestic housing market and its relative importance. The variance estimate of the tax progressive index was applied to the assessment of the impact of several fiscal and tax policies on the fairness of the provincial income level since the reform of tax sharing in China [6].

By constructing the econometric model, the trend characteristics of the statistical index sequence and the intrinsic correlation characteristics of the indicators are studied. For example, in order to measure the size of innovation ability, by building comprehensive indicators, establishing spatial measurement models, researching factors affecting the innovation ability indicators [7,8].Measurement model analysis can depict the complex laws between economic indicators in economic systems, but this method requires strict assumptions, and the outliers in statistical data may affect the test results of multiple collinearity test, unit root test and hetero variance test [9]. Therefore, it is necessary to deal with statistical data and choose a robust parameter estimation method [10]. When dealing with data with rapid changes and large volume like financial data, the estimation results of traditional models will be lack of robustness [11].

The basic idea of frequency domain analysis is to view the time series as the superposition of unrelated periodic components. By studying and comparing the periodic changes of each component, to fully reveal the frequency domain structure of the time series and master its main fluctuation characteristics. Therefore, it has the advantage of the time domain method in studying the periodic fluctuations of time series [12].

2. Related Work

Real Economic Cycle Theory (RBC) regards macroeconomic operations as some combination of long-term growth and short-term fluctuations, distinguishing long-term factors (trends) from short-term factors (fluctuations) when analyzing problems. In practice, RBC theorists have proposed that physics can be used to analyse economic problems [13], and some researchers have used frequency domain analysis to analyse macroeconomic data [14].The sequence data obtained by the statistical index can be decomposed using the frequency domain analysis method into fluctuation components at different frequencies. According to the periodic characteristics presented by the different components, the statistical indicators can be decomposed into trend terms with lower fluctuation frequency, noise terms with higher fluctuation frequency, and periodic terms with fluctuation frequency showing general economic cycle characteristics.

Spectral analysis method provides a means for multi-frequency mixed data analysis. According to the spectral representation theorem, stationary stochastic processes with any covariance can be expressed in the following form:

$$Y_t = \mu + \int_0^{\pi}[\alpha(\omega)\cdot\cos(\omega t) + \delta(\omega\cdot\sin(\omega t)]\,d\omega \quad (1)$$

Where μ is the overall mean, ω means for frequency, $\alpha(\omega)$ and $\sigma(\omega)$ response for random variables. The time sequence data generated using statistical metrics yields the sample spectrum of the data, and then the overall spectrum of spectral estimation. Thus, the frequency representation of the statistical index is obtained.

A filter is a frequency selection component that enables a specific frequency component in the signal through, thus a function to filter a specific frequency. Frequency selection filters are commonly used in the economic field. The HP filters, the BK filters, and the CF filters are most widely used in this field. The HP filter is a high-pass filter that can filter the high-frequency partial data in the signal; the BK filter and the CF filter are a band pass filter and can filter the data of the specified bandwidth. The periodogram uses the squared magnitude of the Fourier series of the input time series data as a measure of the power in the function. Periodogram are an important means of spectral analysis, which are defined as follows:

$$I\left(\frac{j}{n}\right) = \begin{cases} \frac{n}{2}\left({A_j}^2 + {B_j}^2\right), & (\text{n is odd}) \text{ or } (\text{n is even and } j \neq \frac{n}{2}) \\ n(A_j)^2, & j = \frac{n}{2} \end{cases} \quad (2)$$

The frequency $f = j/n$ corresponding to the periodic graph is proportional to the sum of square of the corresponding regression coefficients, so the peak of the periodic graph shows the degree of cosine-sine contribution at different frequencies to the overall index in the overall behaviour of the sequence.

Mutual spectral analysis is also known as cross spectral analysis. After converting the time series to the frequency domain by spectral analysis, we break down each frequency component, compare the periodic changes of each component, and make full use of the spectral density function to get the main frequency component to study the frequency characteristics of the time series, such as the first, lag relationship and consistency of different time series.

The coherent spectral formula is given by the following equation. $f_{xy}(\omega)$ is the cross-spectral density function of the time series x and y which are the normalized mean of the component amplitude product of two sequence frequencies ω. The cross-spectral density is used to measure the absolute correlation of the measured time series on the frequency ω. The value of $c_{xy}(\omega)$ close to the value of 1 indicates the stronger the consistency of the two time series over the frequency ω.

$$c_{xy}(\omega) = \frac{|f_{xy}(\omega)|}{\sqrt{f_x(\omega)f_y(\omega)}}, \quad 0 \le c_{xy}(\omega) \le 1 \quad (3)$$

The formula for the phase spectra is given by the following formula, which reflects the mean of the phase offset of the measured time series at frequency ω. If the value of the phase spectrum is positive, the time series $\{x_t\}$ lags on the frequency ω to $\{y_t\}$; if the phase spectrum is negative, then $\{x_t\}$ leads to $\{y_t\}$ on the frequency ω.

$$\varphi_{xy}(\omega) = \tan^{-1}\left(\frac{q_{xy}(\omega)}{c_{xy}(\omega)}\right) \tag{4}$$

3. Performance evaluation framework of statistical index based on mutual spectral analysis

The performance evaluation method based on the frequency domain analysis decomposes the statistical index into: the trend term, the cycle term and the noise, and compares the trend term and the baseline statistical analysis index to obtain the performance evaluation of the statistical index. Trend term refers to the overall trend in time series data; cycle term refers to the regular content in time series data; noise is the interference item introduced by uncontrollable factors in time series data.

3.1 Data filter selection

Burns and Mitchell et al. have argued that actual periodic macroeconomic fluctuations usually occur in 6-32 quarters [15].Tang's study shows that the macroeconomic data of 2-8 years can use Ravn and Uhlig-adjusted HP filter and CF filter [16].The specific HP filter or the CF filter adjusted with Ravn and Uhlig will need to be selected based on the frequency of the actual data and the sample data size. When the fluctuation wavelength is for 6-8 quarters and the data size is small, we select HP filters adjusted by Ravn and Uhlig. In other cases, we can chose the CF filter.

3.2 Find out the cycle of the time series

The periodogram of the statistical index timing data provides a means to find the frequency information of the data. By analyzing the periodogram, a frequency component with a high contribution to the overall index can be found. Researchers can take component components with high contribution as the main frequency component and select filter types based on the main frequency component.

The intensity components corresponding to the periodic components obtained by the periodogram are the basis of the filter. If the main periodic component corresponding to the periodogram is high frequency, the HP filter adjusted by Ravn and Uhlig was filtered and the filtered intermediate and low frequency sequences were analyzed as trend terms. If the main component of the periodogram is medium frequency data, the CF filter is used for filtering. Take the obtained high frequency data as noise term, conduct perturbation analysis combined with the current situation; medium frequency data as cycle term and low frequency data as trend term.

3.3 Trend terms analyzed by ANOVA

In this paper, we analyse the trend-term polarity of the time series by using ANOVA methods. ANOVA is a statistical method for testing whether the populations have equal means under the same variance conditions. The variance test assumes that different populations satisfy the assumption of homogeneity of variance. Homogeneity of variance test is a method in mathematical statistics to check whether the overall variance of different samples is identical. The rationale is to make some assumption about the characteristics of the population first, and then through statistical inference from sampling studies, making inference [11] about whether this hypothesis should be rejected or accepted. Bartlett's test is commonly used to test the homogeneity of variance of normal distribution populations. The population tested passes the homogeneity of variance test, accepting the assumption that the tested population comes from the same distribution.

3.4 Cycle terms analyzed by mutual spectral analysis

After the multiple time series is converted to the frequency domain by spectral analysis, the corresponding components of the following frequency can be decomposed. The main frequency component can be analyzed by using the spectral density function, and the coherence and leading lag relations are obtained using the coherence spectrum and phase spectrum.

4. Case Study

This paper verifies the effectiveness of this analysis using resident consumer price index (CPI) in China and Alibaba online price index (aSPI) to evaluate the performance of aSPI based on CPI.

4.1 Data about aSPI and CPI

CPI is an authoritative macroeconomic index reflecting the changes in general consumer commodity price and service price level purchased by households. The monthly release cycle, and it is legal statistics of China [5]. aSPI is an all-network online purchase price index formulated by Alibaba to reflect the price level of goods and services on the online shopping platform. It corresponds the smallest leaf category to the "basic classification" of the fixed basket price index. Accordingly, the price index is calculated based on the change of the monthly weighted transaction average price of the sub-category, and the transaction share of the above month is calculated to reflect the change in the overall online shopping expenditure price level of the whole network. It also contains information about general price changes at the commodity level and consumer consumption structure changes under the sub-category.

In this paper, the CPI data [17] from January 2008-April 2016 and the corresponding time periods used the iSPI extended aSPI as the data source. With CPI as the benchmark test term, we test the ability of aSPI to reflect the price changes through the aforementioned statistical index performance evaluation method, and further measure the degree of CPI and aSPI in the price changes reflect time.

4.2 Data preprocessing and stationarity test

To facilitate the analysis of the data, the above data are scaled on the (0, 1) interval and unified data representation. Stationarity means that the statistical features contained in the time series data will not change with time. The premise of using the spectral analysis is that the time-series data meet the stationarity requirements. In this paper, the data were tested for stationarity using the augmented Kiddi-Fuller test (Augmented Dickey-Fuller Test). By testing, we can find that the CPI data during this time period meets the stability assumption; because aSPI only provides relevant data from February 2011 to April 2016, the data cannot pass the stability test, the iSPI data is used to expand the aSPI data. iSPI

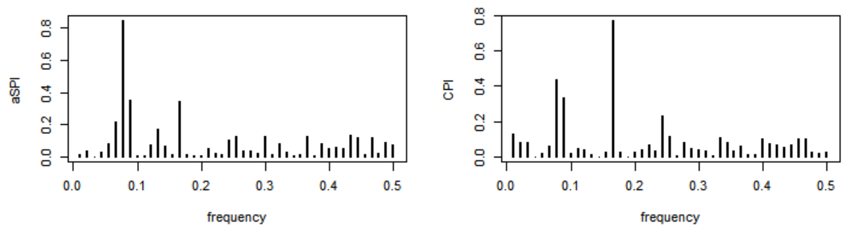


Figure 1. Periodogram of the aSPI and CPIdata

Table 1. Frequencies and Amplitudes of CPI Data

Frequency	0.011	0.022	0.033	0.044	0.056	0.067	0.078	0.089
Amplitude	0.129	0.084	0.08	0.001	0.022	0.064	0.439	0.336
Frequency	0.1	0.111	0.122	0.133	0.144	0.156	0.167	0.178
Amplitude	0.019	0.05	0.043	0.012	0.001	0.027	0.77	0.03
Frequency	0.189	0.2	0.211	0.222	0.233	0.244	0.256	0.267
Amplitude	0.001	0.026	0.037	0.068	0.034	0.232	0.114	0.007
Frequency	0.278	0.289	0.3	0.311	0.322	0.333	0.344	0.356
Amplitude	0.081	0.047	0.042	0.033	0.006	0.111	0.08	0.033
Frequency	0.367	0.378	0.389	0.4	0.411	0.422	0.433	0.444
Amplitude	0.059	0.01	0.016	0.099	0.073	0.065	0.055	0.066
Frequency	0.456	0.467	0.478	0.489	0.5			
Amplitude	0.099	0.1	0.029	0.017	0.025			

This index, also the online shopping price index provided for Alibaba, is a subset of aSPI, and is also calculated using the Roche formula, with the same data source, only with different classification. The ADF test for the new data set after the extended aSPI data resulted that the extended aSPI data satisfies the stationarity assumption of the time series.

4.3 Main cycles analyzed

After profiling of the time series, the time series data can be represented as a combined representation of the corresponding periodic components. The observation of the data cycle diagram shows the contribution of the frequency domain components to the overall index.

Since CPI is the benchmark indicator, the periodic decomposition takes the frequency component affecting CPI as the benchmark component. According to the observation of the periodogram and the analysis of the periodic intensity corresponding to the frequency, excluding the drag tail effect showed the highest intensity at frequency 0.167 and followed at 0.078 and 0.244.Because there are 83 months, the cycle of the strongest component is 14 months, 7 months, and 20 months. After actually calculating the periodic graph of aSPI, the first three strongest components obtained after removing the long tail effect are consistent with the results of the CPI periodic graph analysis.

4.4 Periodic and trend terms selecting

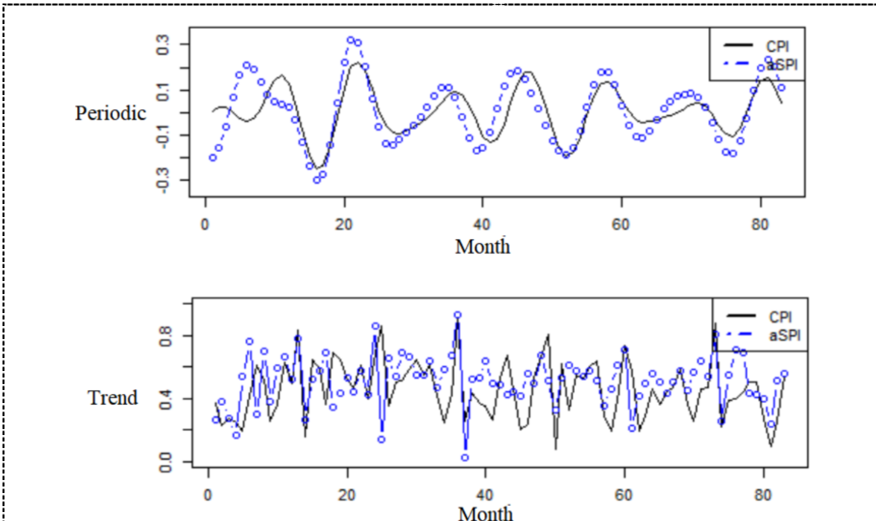


Figure 2. Comparison of periodic terms and trend item between CPI and aSPI

Since the experimental data in this paper satisfy the stability hypothesis, the time series data of CPI and aSPI can be filtered by setting the band pass parameters of the BK filter. According to the results obtained in the above periodiogram, the lower limit of the minimum cycle is 7 months and the upper limit of the maximum cycle for the two time series data is 20 months. The two time series data were filtered to obtain the trend and cycle term data at the corresponding time points. Trend terms and periodic terms show a high degree of consistency in the trends.

4.5 Trend terms were analyzed using ANOVA

For the trend data extracted from the filters, the Bartlett test was used to test the homogeneity of variance of the CPI and aSPI trend term data. Since the null hypothesis of the Bartlett test is the same variance of the overall data tested and the P value of 0.8765, we can conclude that the overall variance of the CPI trend and aSPI term data is the same and the corresponding population comes from the same distribution.

4.6 Mutual spectral analysis

The consistency of the two timing data CPI data and the aSPI data analyzed at different frequencies is described using the coherence data obtained from the mutual spectral analysis. As shown, strong agreement was shown in 0.167, 0.078 and 0.244 in the main frequency components.

Using the phase spectrum analysis, it is concluded that the phase spectrum data in the main components are less than 0, indicating that the CPI data is ahead of the aSPI data in the main frequency component.The experimental results use frequency domain analysis-based methods to verify the strong agreement between aSPI and CPI, and also extend to analyze the speed of aSPI and CPI in reflecting the speed of price change. The results show that CPI can quickly reflect price changes, while the online shopping platform has a certain transmission delay. Therefore, in the experiment with CPI as the

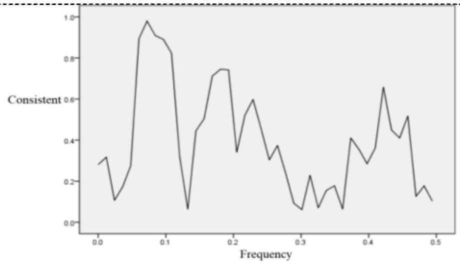


Figure 3. Consistent analysis of the cross-spectral analysis of the CPI and aSPI trend data

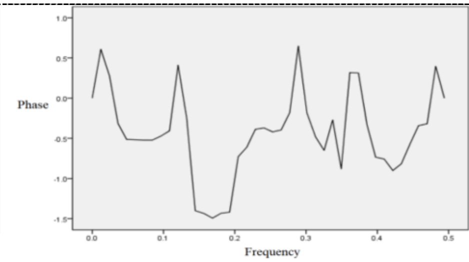


Figure 4. Phase spectrum analysis of the cross-spectral analysis of the CPI and aSPI trend data

benchmark test index, the statistical performance index of aSPI does not reach the performance of CPI.

The emerging analysis framework of statistical index based on spectral analysis still has the following drawbacks in this process:

1. The analysis process adds the calculation of the cycle, trend and cycle consistency and lag relationship, which needs more background knowledge and increase the complexity of the analysis.

2. The choice of filter directly determines the effect of analysis. But the definition of high frequency and intermediate frequency requires researchers to choose the filter according to experience and actual situation. The lack of unified specification limits the application scenarios.

5. Conclusion

This paper proposes a framework based on mutual spectral analysis to evaluate the performance of emerging statistic index compared with the original benchmark indicators. By using this framework, we analyse the index performance of aSPI with CPI data as the benchmark data. With the update of statistical methods, the deepening of new statistical indicators and comprehensive index system, this framework give a system methods to overcome the defects such as insufficient accuracy and strict assumptions, which provides performance evaluation methods for the emerging statistical indicators proposed under the constantly emerging new technical methods.

Acknowledgments

This work is supported the project of Science and Technology Department of Yunnan Province(No.202101AU070158), the projects of Yunnan Education Department (No.2023J0788) and the projects of Yunnan Police College (No.19A021, No.19A010).

References

[1] Xianchun Xu. Comparison of some important indicators in Chinese national economic accounting and relevant statistical indicators .The world economy. 2014, 3(15):145-159.

[2] Jizhe Ning. Statistical exploration and practice of new industries, new business forms and new models. Beijing: China Statistics Press, 2017.

[3] Chenglin Qin, Hua Zhang and Jihui Zhang. New trends and causes of regional development imbalance in China based on the measures of the population-weighted coefficient of variation and its spatial and industrial binary decomposition. China Industrial Economy. 2011(10) :39-47.

[4] Shuang Yao, Weiqiang Huang, Zhan Zhang and Liheng Zhao. Regional Financial Ecological Environment Evaluation: based on the modified coefficient of variation method. Technical Economy, 2015, 34(10):61-67.
[5] Xiuhai Huang, Comely Teng, Dan Zhao. Comparative Study of Online Shopping Price Index and CPI. Statistics and Consulting. 2017(6):25-28.
[6] Shengping Dai. Variance estimation of the tax-progressive M index and its application. Quantitative, economic, technical and economic research. 2019, 36(3): 124-136.
[7] Yi Su, Zhou Lin. Study on the spatial effects and influencing factors of regional innovation activities. Quantitative, economic, technical and economic research. 2017, 34(11):63-80.
[8] Xiaohua Jin, Taihua Yuan. International Technology Overflow and Regional Innovation Capacity: Empirical Analysis from an IPR protection perspective. International Trade Issues. 2017(3):14-25.
[9] Jinwen Zhao. Typical cases of the effect of outliers on measurement modeling. Statistical study. 2010, 27(12):92-98.
[10] Zhijian Wang. Construction and comparative study of robust AR models. The Practice and Knowledge of Mathematics. 2019, 49(13) :156-166.
[11] Zongxin Zhang, Xuejiao Zhang. Fund trading behavior and market volatility: based on the wavelet and mutual spectrum analysis of data mining. Journal of Management Engineering. 2012, 26(2):156-161.
[12] Hengzhen Lu, Liping Xu. Futures Market Cycles Based on Time-series frequency domain analysis. Statistics and Decision Making. 2011(6):146-147.
[13] Prescott K. The Computational Experiment: An Econometric Tool. Journal of Economic Perspectives. 1996,10(1):69-85.
[14] Baxter M, King R G. Measuring Business Cycles: Approximate Band-Pass Filters For Economic Time Series. Review of Economics & Statistics. 1999, 81(4):575-593.
[15] AF Burns, Mitchell W C . Measuring Business Cycles. Nber Books. 1946, 78(1) :67-77.
[16] Jinding Tian, Qiangzhong Zhu, Fengxia Zhang, Junling Song. Application of ANOVA in product testing. Journal of Liaocheng University (Natural Science Edition). 2008, 21(1):23-24.
[17] Xianglan Yu. Empirical test of the effectiveness of macro-statistical index data in China. Statistics and Decision Making. 2011(23):8-12.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230191

Attitudes Towards Framed Communication of Political Parties in North Macedonia During Second Wave COVID 19 Through the Lenses of Supporters

Ljupcho Efremov[1]
American University of the Middle East, Kuwait

Abstract. The appearance of COVID19 triggered governments to start managing the crises through communicating centers to influence the perception of public. One part of crisis management was through political communication. The aim of this paper is to explore whether there are differences among opposing party members in perception of political communication in times of crises. Media headlines from both parties were evaluated about degree to agreeableness with each statement by the members. The sample consisted of members of two largest political parties from the Macedonian bloc VMRO-DPMNE and SDSM (50 members from each party). The analysis was done through t-test comparison of attitudes towards message between party members. The results show that each party member supports their own party framed communication. Main conclusion is that there is a difference in perception of political communication related to COVID-19 which is in favor of own political affiliation. The findings provide a solid base for crisis management through political communication in uncertain situations.

Keywords. Framed communication; political communication; Covid-19; party membership; supporters

1. Introduction

Beliefs about Covid-19 are regarded complex as there have been different public reactions to understanding the virus. One part accepted the announcements from the WHO and the authorities, while another part showed disbelief in the existence of the virus, implying problem awareness. Political parties, especially those in power, needed to set up references. The purpose was to ease the interpretation presented to the public [1]. The existence of COVID-19 enabled political parties to play with public sentiment even more easily because the public did not not know how to choose relevant benchmarks in this new situation. Previous research has found that political leaders can influence how seriously people take a problem and this is a case for party members or supporters [2], [3]. Scholars already verified that the public instinctively uses heuristics to reduce the burden of decision-making [4], especially on issues like COVID-19, where

[1] Corresponding Author, Ljupcho Efremov, American University of the Middle East, str 250 Egaila Kuwait E-mail: Ljupcho.efremov@aum.edu.kw

there is an apparent disconnect between scientific understanding and mass competence [5]. However, those signals are often unreliable and prone to cognitive bias. When voters act on signals given by their preferred political party and selective exposure to preferred media or preferred sources [6], they may do so automatically.

"The pictures in our heads", to borrow Walter Lippmann's famous phrase [7] are shaped less by factual knowledge and more by various other factors that are not easily controlled, such as personal experience and chosen cues from the real world. People's perception of risk (in this case of COVID-19) is not only determined by the scientific information they receive or their physical experiences. Vai et al. [8] points to the media that shape public risk perception as a critical factor influencing risk perception. They also highlight the different types of media factors that influence public risk perception, such as the type of media, the amount and tone of coverage, and the credibility of the source. More specific to this research are political identities that also take part in cognitive factors towards risk perception [1]. As political identities shape components of cognition, political psychology offers irrefutable evidence that partisan identities alter political thinking and reasoning [9] on political cognition [10] and social themes [11]. Therefore, the aim of this research is to explore whether there are differences among opposing party members in perception of political communication in times of COVID-19 crises.

2. Influence of political affiliation

Opposing views of political polarization maintain influence on political behavior within developed countries and developing democracies [12]. In Republic of N. Macedonia, political membership is large and highly polarized [13] and tends to imply an impressive influence on the range of political behavior. According to some data, Jovanovska and Božinovska claim that 13 percent of citizens in Macedonia are members of one of the three largest parties: VMRO-DPMNE, SDSM or DUI.

Party membership represents a form of group identity that shapes the behavior of the individual towards the group and outside it. The polarization in relation to the opponents increases through the introduction of strong rhetoric and is supported by media [14]. Thus, a polarized political climate is characterized by “a high level of ideological distance between parties and a high level of homogeneity within parties” [15]. Certain authors [16] point to the fact that political parties in Macedonia constantly use polarized discourse as a populist strategy that builds the basis of division more thoroughly. The theory of political polarization proves that as the opposition of political parties increases over a certain issue, their respective ideological identification designations will become more distinct and thus more easily recognized by their members [17]. Attitudes of parties constantly seek to occupy the importance of a certain issue or problem [18], and this creates additional impact on polarized public opinion. However, if there is unification of party attitudes, as the Canadian example [19] (usage of unifying discourse and overcome party short term interests), then political differences are reduced and there is some cooperation. The specific public that recognizes the source of the message will be willing to interpret the message with party perspective [20]. As some authors [21] note, party cues may not indicate the character of politics, but they do indicate party groups that are important. Through such behavior the parties tend to encourage the public to judge the ultimate impressions through an ideological prism of politics [21]. The findings relate to a study that found [22] influence of emotions in organizational context. Also, it was [23] found that exposure to candidate rumors is positively related to belief in said rumors for members of both parties, but the relationship is significantly stronger when the rumors are attitudinally congruent. Thus [24] it is demonstrated that rumors are more effectively corrected by subjects who argue against their political interests than by sources who might be expected to oppose the content. Research shows [25] that subjects tended to distrust scientific evidence when presented as subject to party disputes. A review of research on this topic finds that party signals in media coverage about politics contribute to polarization that increases the visibility of party views. Politically biased media have been extensively studied. In one [15] experiment, which was conducted by manipulating the types of arguments (weak and strong arguments), as well as information about the level of opposition between two opposing political parties, the results showed a trend toward following stronger arguments, but also the consistency of the argument, regardless of its strength if it shows the party. The abovementioned provide a basis for the problem of interest of this study. The main hypothesis is that there is a difference in perception of political communication related to COVID-19 which is in favor of own political affiliation.

3. Methodology

The research uses a convenient sample composed of 50 members of the opposition party VMRO-DPMNE and 50 members of the ruling party SDSM who are older than 18 years. The data was collected through a questionnaire placed on Google Forms. The study was conducted in the period of March 20, 2021, to June 5, 2021. The total number of respondents was 100 (50% males and 50 % females from both subsamples of members). All variables, except demographics, were measured using a self-evaluation scale corresponding to a Likert scale. The questions are evaluation of political statements given by members of parties in the media. The questionnaire consisted of an evaluation of each statement (media headline) on a 1 to 5 scale where 1 = strongly disagree and 5 = strongly agree. This approach was used to measure the attitudes towards politicians from both political parties. The participants of both parties were shown media titles from their own and opposing party. This included a total of 18 media titles, i.e. 9 of SDSM and 9 of VMRO-DPMNE. To test the proposed hypothesis, a T-test for independent groups was used e.g., the analysis consisted of comparison of attitudes between members of opposing parties for own media headlines and for media headlines of opposing party.

4. Results

The data from the analyzed survey questionnaire shows significant results for each of the variables according to the hypothesis. For this purpose, two tables for each of the messages are presented with comparison of attitudes of members of both parties.

Table 1. Comparison of attitudes between members of SDSM party and members of VMRO-DPMNE party towards quotes with framed messages by VMRO-DPMNE

Aspects	Average of SDSM members	SD	Average of VMRO-DPMNE members	SD	P value
1. VMRO-DPMNE: Zaev and Filipche let the situation with the pandemic out of control a long time ago (24 TV, 05.10.2020)	1.44	0.20	3.84	0.20	**p<0.05**
2. VMRO-DPMNE: Citizens suffer from the government's inability to deal with Covid-19 (Plusinfo, 18.10.2020)	1.42	0.19	3.64	0.14	**p<0.05**
3. VMRO-DPMNE: Every fourth person tested is infected, the government is doing nothing to prevent an even bigger disaster (Plusinfo, 20.10.2020)	1.53	0.19	3.72	0.17	**p<0.05**
4. VMRO DPMNE: The government should overcome vanity and elect a new Commission for Infectious Diseases (Plusinfo, 23.10.2020)	1.8	0.19	3.7	0.17	**p<0.05**
5. VMRO-DPMNE demands the resignation of Minister Filipce (Focus, 27.10.2020)	1.41	0.22	2.51	0.16	**p<0.05**
6. VMRO-DPMNE: The government of SDSM and Zaev is not capable of managing the economic and health crisis (Channel 5, 15.11.2020)	1.69	0.18	3.23	0.18	**p<0.05**
7. VMRO-DPMNE: Filipce is the biggest problem for the corona crisis in the country, resignation immediately (Sitel, 20.11. 2020)	1.38	0.22	3	0.14	**p<0.05**
8. VMRO-DPMNE ACCUSES: The measures are a lie and PR (Sitel, 15.12.2020)	1.65	0.21	2.3	0.19	**p<0.05**
9. VMRO-DPMNE: The peak of the corona virus does not pass as well as the incompetence that means death for the people (A1 ON, 04.12.2020)	1.81	0.27	3.38	0.20	**p<0.05**

The messages sent by the political party VMRO-DPMNE, shown in table 1, offer evidence ($p<0.05$) in all 9 cases that members of the opposite party or SDSM do not agree with their contents (message) and are supported by own members (VMRO-DPMNE).

Table 2. Comparison of attitudes between members of SDSM party and members of VMRO-DPMNE party towards quotes with framed messages by SDSM

Aspects	Average of SDSM members	SD	Average of VMRO-DPMNE members	SD	P value
1. Filipche accused VMRO-DPMNE of blocking: The virus does not wait, the measures must be adopted urgently! (Focus, 21.10.2020)	3.79	0.23	2.00	0.15	**p<0.05**
2. SDSM: If VMRO-DPMNE had not requested a postponement, the measures would have been passed in the Parliament (Nova TV, 22.10.2020)	3.87	0.23	1.76	0.16	**p<0.05**

3. SDSM: Since the beginning of the pandemic, VMRO-DPMNE has been holding back all efforts to deal with the crisis (Focus, 31.10.2020)	2	0.18	1.5	0.21	**p<0.05**
4. SDSM: Mickoski remains on the side of the coronavirus and acts dangerously for the health of citizens (Nova TV, 09.11.2020)	3.62	0.18	1.4	0.21	**p<0.05**
5. SDSM: VMRO-DPMNE blocked funds for 16 covid-centers, reducing the victims of coronavirus to numbers (MIA, 16.11.2020)	4.4	0.17	1.83	0.22	**p<0.05**
6. SDSM: The hypocritical VMRO-DPMNE and Mickoski feign concern (Channel 5, 16.11.2020)	3.83	0.19	1.7	0.20	**p<0.05**
7. SDSM: VMRO-DPMNE is abusing the misfortune of citizens infected with the coronavirus (MIA 17.11.2020)	3.7	0.17	1.5	0.17	**p<0.05**
8. SDSM: Mickoski secretly rejoices at the number of deceased (Sitel, 19.11.2020)	3.25	0.18	1.34	0.22	**p<0.05**
9. SDSM: VMRO-DPMNE feigns concern and violates measures to protect against Covid (24 TV, 17.12.2020)	3.73	0.20	1.9	0.20	**p<0.05**

The messages sent by the political party SDSM, shown in table 2, represent a difference ($p<0.05$) in all 9 cases for motivated party reasoning, where members of the same party SDSM support the statements. Members from the opposition party VMRO-DPMNE disagree with the messages.

5. Discussion

The overall result is that each party member supports their own party framed communication e.g., when it comes to the media quotes whose message framing originates from the opposing party, there is a difference in supporter's attitudes. Each party member supports the framed communication of the party they favor, while at the same time highly oppose and disagree with framed communication of the opposing party. Mostly, the accumulated data using the framed messages in both groups indicates that the views of the political members are in line with the intention of the message to provoke opposition [26]. The conclusion is that there is a difference in perception of political communication related to COVID-19 which is in favor of own political affiliation.

Accordingly, the research confirmed the polarized attitudes among members of political parties in a communication environment where fragmented political information encourages and maintains polarization [26]. The research supported the thesis about the existence of differences in attitudes toward framed communication of politicians who are in the position between members of the party in power and members of the party in opposition. Therefore, political parties should take greater steps regarding the strategic management of political information. The results from this study are in line with research that shows how information obtained from different sources significantly influences the formation of public attitudes and beliefs [27]. The study contributes to the effect of polarization e.g., the tendency of the opposition not to support the policy that is proposed by those in position, and vice versa [18], as was the case in this study. The abovementioned finding is interpreted through the prism of politicization of existing opinion where political members accept and support the position that originates from the elites of their preferred political party. As such, information obtained from political sources not only provides information about the surrounding world, but also enables the interpretation of that information [28]. According to some authors [29], political members are motivated to protect their identity and choose its continuity, thus becoming more motivated to view the competing party in a more negative light.

6. Conclusions

The conclusions that stem out from the research are that there is a highly polarized climate which is characterized by usage of polarized discourse as a populist strategy. Again, the theory of political polarization proves that as the opposition of political parties increases over a certain issue, their own affiliations play a role when it comes to attitudes. source. The results are due to the political identities of these specific groups and their affiliation which is fed by polarized climate.

The findings are significant for political communication and communication in crisis by the state. The conclusions point out that members of political parties need to take polarization into account when framing the messages in times of crises. This can be achieved through using cognitive arguments that sound reasonable to members of both parties. Also, by using ideological cues and frames, political parties can fundamentally change the way that the same public perceives various issues. On the other hand, both parties need to use a unifying discourse and overcome party short term interests in front of public interests. Regarding limitations, the study uses convenience sampling and due to this selection method, the results should be taken with some caution. A change in the reasoning and motivation of such behaviors may occur due to different reasons and therefore further research into potential variations is necessary. However, the obtained data provides nuances of knowledge about effects of political communication and its impact on opinion formation in risky environments.

References

[1] Van Bavel JJ, Pereira A. The partisan brain: An identity-based model of political belief. Trends in cognitive sciences. 2018 Mar 1;22(3):213-24.

[2] Bisgaard M, Slothuus R. Partisan elites as culprits? How party cues shape partisan perceptual gaps. American Journal of Political Science. 2018 Apr;62(2):456-69.

[3] Barrios JM, Hochberg Y. Risk perception through the lens of politics in the time of the COVID-19 pandemic. National Bureau of Economic Research; 2020 Apr 20.

[4] Bolsen T, Druckman JN, Cook FL. The influence of partisan motivated reasoning on public opinion. Political Behavior. 2014 Jun;36:235-62

[5] Kraft PW, Lodge M, Taber CS. Why people "don't trust the evidence" motivated reasoning and scientific beliefs. The ANNALS of the American Academy of political and social science. 2015 Mar;658(1):121-33.

[6] Stroud, N.J., Muddiman, A. and Scacco, J.M., 2017. Like, recommend, or respect? Altering political behavior in news comment sections. New media & society, 19(11), pp.1727-1743.

[7] Coleman R. Framing the pictures in our heads: Exploring the framing and agenda-setting effects of visual images. InDoing news framing analysis 2010 Feb 26 (pp. 249-278). Routledge.

[8] Vai B, Cazzetta S, Ghiglino D, Parenti L, Saibene G, Toti M, Verga C, Wykowska A, Benedetti F. Risk perception and media in shaping protective behaviors: insights from the early phase of COVID-19 Italian outbreak. Frontiers in psychology. 2020 Nov 5;11:563426.

[9] Boyer MM, Aaldering L, Lecheler S. Motivated reasoning in identity politics: Group status as a moderator of political motivations. Political Studies. 2022 May;70(2):385-401.

[10] Leeper TJ, Slothuus R. Political parties, motivated reasoning, and public opinion formation. Political Psychology. 2014 Feb;35:129-56.
[11] Petersen MB, Skov M, Serritzlew S, Ramsøy T. Motivated reasoning and political parties: Evidence for increased processing in the face of party cues. Political Behavior. 2013 Dec;35:831-54.
[12] Bavel JJ, Baicker K, Boggio PS, Capraro V, Cichocka A, Cikara M, Crockett MJ, Crum AJ, Douglas KM, Druckman JN, Drury J. Using social and behavioural science to support COVID-19 pandemic response. Nature human behaviour. 2020 May;4(5):460-71.
[13] Jovanovska, S. Božinovska, Z, P. (2017). Komparativna studija za izbornite nedostatoci vo Makedonija i iskustvata vo regionot. Centar za graganski monitoring. Magnasken, Skopje.
[14] STOA (2019). Polarisation and the Use of Technology in Political Campaigns and Communication. European Parliament Research Service.
[15] Druckman JN, Peterson E, Slothuus R. How elite partisan polarization affects public opinion formation. American Political Science Review. 2013 Feb;107(1):57-79.
[16] Blaževa, A., Kolozova, K., Demiri, M., Saveski, Z. Поларизирачкиот дискурс и влијанието врз политичката и социјалната поларизација во македонското општество. Институт за општествени и хуманистички науки, Скопје. 2019.(In English: Polarized discourse and influence of political and social polarization in Macedonian society. Institute for Social and Humanistic Sciences, Skopje. 2019.)
[17] Nicholson SP. Polarizing cues. American journal of political science. 2012 Jan;56(1):52-66.
[18] Mullinix KJ. Partisanship and preference formation: Competing motivations, elite polarization, and issue importance. Political Behavior. 2016 Jun;38:383-411.
[19] Merkley E, Bridgman A, Loewen PJ, Owen T, Ruths D, Zhilin O. A rare moment of cross-partisan consensus: Elite and public response to the COVID-19 pandemic in Canada. Canadian Journal of Political Science/Revue canadienne de science politique. 2020 Jun;53(2):311-8.
[20] Slothuus R, De Vreese CH. Political parties, motivated reasoning, and issue framing effects. The Journal of Politics. 2010 Jul;72(3):630-45.
[21] Suhay E, Grofman B, Trechsel AH, editors. The Oxford handbook of electoral persuasion. Oxford University Press, USA; 2020.
[22] Efremov L. Emotions and Attitudes Towards Safety—Relationship Between Affective Commitment and Safety Attitudes Among Construction Employees in North Macedonia. In Occupational and Environmental Safety and Health IV 2022 Sep 17 (pp. 395-407). Cham: Springer International Publishing.
[23] Weeks BE, Garrett RK. Electoral consequences of political rumors: Motivated reasoning, candidate rumors, and vote choice during the 2008 US presidential election. International Journal of Public Opinion Research. 2014 Dec 1;26(4):401-22.
[24] Berinsky AJ. Rumors and health care reform: Experiments in political misinformation. British journal of political science. 2017 Apr;47(2):241-62.
[25] Bolsen T, Druckman JN. Counteracting the politicization of science. Journal of Communication. 2015 Oct 1;65(5):745-69.
[26] Yeo SK, Cacciatore MA, Scheufele DA. News selectivity and beyond: Motivated reasoning in a changing media environment. Publizistik und gesellschaftliche Verantwortung: Festschrift für Wolfgang Donsbach. 2015:83-104.
[27] Esser F, Hanitzsch T, editors. Handbook of comparative communication research. New York: Routledge; 2012 Apr.
[28] McNair B. An introduction to political communication. Taylor & Francis; 2017 Jul 6.
[29] Groenendyk E. Justifying party identification: A case of identifying with the “lesser of two evils”. Political Behavior. 2012 Sep;34:453-75.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230192

The Digital Intelligent Application of Experiment Teaching Based on SVM-Adaboost Algorithm

Jianghui Liu [a, 1], Runlong Wu [b]
[a] *Modern Education Technology Center, Experimental Teaching Center, Guangdong University of Foreign Studies, Guangzhou, China.*
[b] *School of Finance, Guangdong University of Foreign Studies, Guangzhou, China.*

Abstract. This study applied support vector machine algorithm and adaptive-boost algorithm to analyze the best division hyperplane of enterprise resource planning experimental teaching. We used two groups of experimental data to apply support vector machine and adaptive-boost algorithm. To complete data preprocessing and assign different weights of each index we applied adaptive-boost algorithm. Then we used the SVM to calculate and classify the expected samples. After two sets of experiments, the results show that the expected samples classified by support vector machine and adaptive-boost algorithm have a better fit with the actual experimental situation. It means that the algorithm improves the ability of digital intelligent prediction and feedback in experimental teaching. It supplies a reference for the experimental teaching of the immersive economy and management major in the future.

Keywords. SVM algorithm; Adaboost algorithm; digital intelligent; experimental teaching

1. Introduction

In modern society, emerging information technologies such as 5G, digitalization and intelligence are developing at a rapid pace. We should maintain new momentum and new advantages in the digital technological revolution and industrial transformation [1]. We should strengthen the development of science and technology in human resources and personnel training, and implement digitalization and intelligence in these two aspects [2]. The trend of digital intelligence represents the trend of digital intelligence in teaching mode. The use of digital intelligent information technology makes the teaching flow with the tide, the teaching level is higher and the teaching method is more efficient [3]. Building an immersive digital intelligence laboratory can enhance the educational application of digital intelligence [4]. This is great significance to the major of economic management under the background of new liberal arts [5]. This study will be based on SVM-Adaboost algorithm. The aim of this study was to experimentally determine the digital intelligent function of the algorithm model in ERP experimental teaching.

[1] Corresponding Author: Jianghui Liu, Modern Education Technology Center, Experimental Teaching Center, Guangdong University of Foreign Studies, China; This study is one of the research achievements of the Guangdong Higher Education Teaching Research and Reform Project in 2022. Email: 247031690@qq.com

2. Background of SVM combined with other algorithm models

Souza, F et al employed SVM to predict risks in general surgery, and showed in the process of optimization and experiment that SVM model can accurately predict different types of risks, as well as the causes and measurement methods of risks [6]. Vazquez, J et al utilized SVM supervised machine learning classifiers to estimate de-identification and efficient classification of clinical trials [7]. Zhang A et al studied the short-term load of the micro-grid based on the DA-SVM algorithm model, and achieved accurate prediction through SVM, which improved the operation economy of the entire micro-grid [8]. Zhong L et al used the GPS/INS combined navigation soft fault detection method based on LS-SVM and AIME to classify the soft fault problems in navigation process and the causes. The efficiency of soft fault detection is strengthened by SVM learning. It is proved that it has good ability in intelligent prediction and processing of performance evaluation data. Some scholars have studied the experimental comparison of two models based on SVM algorithm and BP neural network algorithm. The results show that the SVM algorithm model is superior and efficient in the case of small samples and big data gray scale. The background of this research is the feedback of teaching effect of numerical intelligence experiment. The overall sample size was small. Therefore, support vector machine algorithm model is the first choice in this study [9]. Combining support vector machine algorithm with other algorithms can make up the deficiency of support vector machine and improve the actual efficiency of research. In this study, we combine SVM algorithm with Adaboost (Adaptive Enhancement) algorithm. Combined with other algorithms, the algorithm model can deal with many kinds of small samples and complex content well. In this study, considering the actual situation of the research object, we will use SVM-Adaboost algorithm to solve problem and do research.

3. SVM-Adaboost algorithm

3. 1 Adaboost algorithm

Adaboost is an iterative machine learning algorithm. It can be used as a basic classifier in combination with other algorithms. What's more, obtain different classifiers after training a sample data set. Generally, it constructs a linear combination of weak classifiers, and then assembles them to form a stronger classifier. After that, the weight of sample distribution in the sample data set is analyzed to form a strong classifier with better model performance. Through the iterative analysis of the "discourse power" among different samples, the higher the discourse power is, the higher the weight will be assigned in the next iteration. Basic weak classifiers with higher accuracy can obtain greater weight in the end. A strong classifier formed by its combination will have a better effect. The basic idea of the algorithm is as follows.

1. Initialize the weights of the sample data set: $\rho_1 = \frac{1}{N}$, the total number of samples is denoted by N, the weight of the first iteration sample is expressed as ρ_1.

2. After the number $e = 1,2, \cdots, E$ of iterations, use the sample data set with the existing weight ρ_e. After obtaining a certain number E of iterations, the basic weak classifier in the first iteration is obtained g_e.

3. Calculate the classification error of g_e. Get the classification accuracy s_e.

4. Based on the calculated classification accuracy and the basic weak classifier, the weight coefficient expression is the formula.

$$\varepsilon_e = \frac{1}{2} log \frac{1-s_e}{s_e} \tag{1}$$

5. The weights are updated again according to the classification results. The expression of linear combination of basic weak classifiers is the formula.

$$W \ (x) \ = \sum_{e=1}^{E} \varepsilon_e \cdot g_e \ (x) \tag{2}$$

6. The resulting strong classifier combination expression is the formula.

$$J \ (x) \ = sign\left(\sum_{e=1}^{E} \varepsilon_e \cdot g_e \ (x) \ \right) \tag{3}$$

Firstly, the Adaboost algorithm is used to preprocess our experimental data, that is, the weight of the whole data is pre-allocated. At the same time, the error rate of each weak classifier is calculated, and finally the weak classifier with small error rate is combined into a strong classifier. This strong classifier is then combined with SVM algorithm to complete the whole binary classifier.

3. 2 SVM algorithm

SVM algorithm is widely used in data regression and classification. As a binary model, its basic idea is to map the feature vector of an instance into a vector of points in the space. Then draw a line to distinguish the two types of points "as best as possible" and keep a kind of feature space after distinguishing them. So that the new points that appear later can also be distinguished by this line and classified in the appropriate and exact feature space. In general, we can call this line a partition hyperplane, and theoretically there can be an infinite number of partition hyperplanes in a sample space. We want to find the partition hyperplane that is "as good as possible". As shown below, it is obvious that the thicker line in the middle is the optimal partition hyperplane we are looking for.

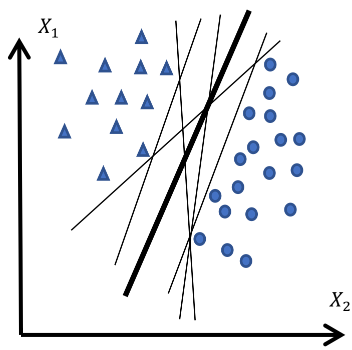


Figure.1 Classification of two types of samples by multiple partition hyperplanes

The equation of such a hyperplane can be expressed by the following formula:

$$T^h m + u = 0 \tag{4}$$

If there are only two types of samples in the sample space, then there is always a minimum Euclidean distance between the sample points on both sides and the same hyperplane of the two samples. This distance is the two sample points closest to the same hyperplane. The minimum Euclidean distance can be drawn in different ways and the hyperplane is just trying to draw it as far away as possible. Greater spacing between sample points means fewer errors and better robustness. Therefore, in the figure above, we choose the thicker partition hyperplane in the middle as the optimal partition hyperplane.

So, the key of using SVM algorithm to solve the problem is to find the optimal partition hyperplane. The two sample data sets are set as $(m_i, r_i), i=1, 2, \cdots n$. We use $T^h m + u = 0$ to express partition hyperplane. The constant u can be understood as the intercept in the equation of the line. T is the adjustable weight vector. They are plane function weights. The h in the formula (4) is the transpose of the vector. In this case, the constraint conditions to be satisfied by the hyperplane are the formula.

$$r_i[(T \cdot m_i) + u] \geq 1 \tag{5}$$

In order to maximize the accuracy of sample classification, we need to ensure that the selected partitioned hyperplane has the best generalization ability. That is, we want to compute a boundary with the greatest distance in all hyperplanes. Since the hyperplane corresponds to the margin, it is necessary to find the maximum margin to find the best generalization ability of the hyperplane.

Determining T、u to maximize the margin, then we have the following formula.

$$arg \underset{T,\ u}{max} = \left\{min\left(r(T^h + u)\right) \cdot \frac{1}{\|T\|}\right\} \tag{6}$$

For the convenience of calculation, equivalently express the formula (6) as follows.

$$min \frac{1}{2} \|T\|^2 \tag{7}$$

The norm of dividing hyperplanes is expressed by $\|T\|$。

Then we use Lagrange multiplier method to solve formula (7), and the optimized expression is shown as follows.

$$L(T,\ u,\ \beta) = \frac{1}{2}\|T\| - \sum_{i=1}^{n} \beta_i \left(r_i\left((T \cdot m) + u\right) - 1\right) \tag{8}$$

The L represents the transformation operator, β is a weight used in the constructor. We take the partial derivative with respect to $T,\ u$. Get a function of β, and turn that function into an equivalent formula.

$$max(\beta) = \sum_{j=1}^{n} \beta_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \beta_i \cdot \beta_j \cdot r_i \cdot r_j\left(m_i^h \cdot r_j\right)$$
$$\sum_{j=1}^{n} \beta_i r_i = 0 \tag{9}$$

The optimal parameter calculation formula are as follows.

$$T_{best} = \sum_{j=1}^{n} \beta_i r_i m_i \tag{10}$$

$$u_{best} = r_i - \sum_{j=1}^{n} r_j \beta_j \left(m_j \cdot m_i\right) \tag{11}$$

The optimal classification function can be obtained according to the derivation. Now we can calculate this optimal classification formula through the previous calculation process and code package, and then realize the prediction of data through this formula.

$$\begin{aligned} f(x) &= sgn[(T_{best} \cdot m) + u_{best}] \\ &= sgn\{[\sum \beta_j r_j (m_j \cdot m_i)] + u_{best}\},\ m \in R^n \end{aligned} \tag{12}$$

Now the problem seems to have been solved "as far as possible". However, to find such an optimal hyperplane and completely divide the two types of samples requires a prerequisite that the sample data itself is completely classifiable, with no indivisible parts. But that's obviously not possible. So, we also need to consider the case of soft margin.

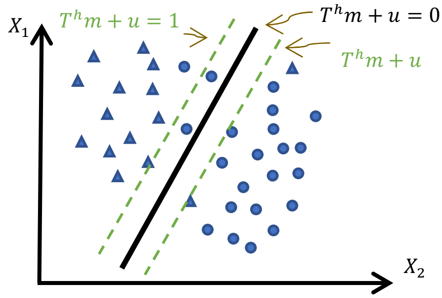


Figure 2. soft margin Diagram

Obviously, there are some samples that don't satisfy the hyperplane constraint. At this point we add a relaxation variable to give the loss of the classification error, in terms of τ. It can be expressed by a common loss function, such as MSE, Sigmoid, etc. However, based on the applicability of SVM algorithm, we used another Loss function, Hinge Loss to stand for τ_i. Then the original constraint will change as follows.

$$r(T^h m_i + u) \geq 1 - \tau_i, \quad i = 1,2 \cdots n \tag{13}$$

The τ_i expresses the deviation margin allowed for i^{th} sample point. We need to minimize $\frac{1}{2}\|T\|^2$ to minimize the classification error. So, its formula is as follows.

$$min = \frac{1}{2}\|T\|^2 + P\sum_{i=1}^{n}\tau_i \tag{14}$$

And we add the penalty coefficient P. The more correct the classification model is, the greater the penalty coefficient will be. Now we have a new Lagrange formula.

$$L(T,\ u,\ \tau_i) = \frac{1}{2}\|T\|^2 + P\sum_{i=1}^{n}\tau_i - \sum_{i=1}^{n}\beta_i\left[r_i\left(T^h \cdot m_i + u\right) - 1 + \tau_i\right] - \sum_{i=1}^{n}\tau_i \tag{15}$$

Now we can add relaxation variables and SVM can solve the problem of data confusion or sample bias. We can vary the penalty coefficient to get different results. The specific value can also be adjusted according to the actual problem to be solved by the research object. This is the linearly separable case we described above.

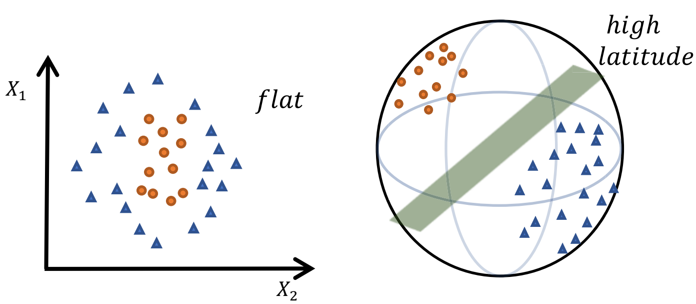


Figure 3 Search for optimal hyperplane by entering high latitude model.

For example, when two types of samples meet this situation, it is linear indivisible. It is impossible to select a good line for effective division under the plane display. It means that the plane needs to be expanded and dimensional zed. Therefore, we can find the optimal hyperplane in the high-dimensional model for classification, so that the linear indivisible problem will be transformed into a linear separable problem. In SVM, we can use the kernel function to map the map on the plane to the higher dimensional space to get a linear solution. There are many common kernel functions: linear kernel, polynomial kernel, Gaussian kernel and so on. Here we use the Gaussian kernel as our kernel function. Its formula is as follows.

$$K(x_i, x_j) = exp\left(-\frac{\|x_i - x_j\|^2}{2\delta^2}\right) \tag{16}$$

The δ represents broadband. Based on the expression for the kernel, we have the expression for the target formula.

$$max = \sum_{j=1}^{n} \beta_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \beta_i \cdot \beta_j \cdot r_i \cdot r_j K(x_i, x_j) \tag{17}$$

$$And\; 0 \le \beta_i \le P, \ \sum_{j=1}^{n} \beta_i r_i = 0.$$

The final optimal classification function expression is as follows.

$$f(x) = sgn\left[\sum_{j=1}^{n} \beta_i r_i K(x_i, x_j) + u\right] \tag{18}$$

4. The specific application of the algorithm and experiment teaching

We took advantage of Python programming software for experimental analysis, obtained sklearn. Adaboost Classifier from sklearn library, and downloaded SVM algorithm module from the library. Finally, the data generated by the experiment are trained after preliminary processing. The algorithm steps of this study are as follows.

Step 1: Put two types of sample data sets into the Adaboost algorithm for iterative classification of sample weights. Combine the strong classifier by Adaboost with SVM algorithm。 Use the combined algorithm to process data.

Step 2: Enter the sample $Z= (m_i, r_i),\ i=1, 2, \cdots n$. Execute the SVM-Adaboost algorithm.

Step 3: To find the hyperplane that can best divide the two types of data samples. It is satisfied to $r(T^h m_i + u) - 1 \ge 0$. Calculate the target function $min\frac{1}{2}\|T\|^2$.

Step 3: Determine whether the sample data can be classified well under general constraints. If not, add relaxation variable τ_i to soften the maximum margin between the maximum support vectors. It can allow less classification errors to occur.

Step 4: Determine whether the classification of the sample data set is linearly separable or linearly indivisible. We use different Lagrange functions and optimal classification functions to solve different situations.

Step 5: Solve the optimal parameters of the objective function $T_{best} = \sum_{j=1}^{n} \beta_i r_i m_i$ and $u_{best} = r_i - \sum_{j=1}^{n} r_j \beta_j (m_j \cdot m_i)$ to obtain the optimal classification model. Get corresponding optimal hyperplane expression $f(x) = sgn[(T_{best} \cdot m) + u_{best}]$.

Step 6: Plug in the new sample data set Z^+. Use the optimal hyperplane expression to solve. Get the result and judge the classification.

On the basis of the algorithm steps, we carry out experiments, we found 36 students and divided them into a group of 18 students for two groups to participate the digital intelligent ERP experiment teaching.

The results show that the feasibility of the SVM-Adaboost algorithm model is verified through the comparative analysis of the two groups of experiments. After calculating the experimental data of the first group of experiments by the algorithm, it is proved that the data prediction can be made according to the formula obtained by the algorithm. The prediction and results of the second group of experiments show that the SVM-Adaboost algorithm model successfully predicts the correct experimental teaching method with numerical intelligence. Teachers used the algorithm to calculate the results of the predicted samples, adjust the teaching method, content and details in real time, and process and feedback the actual data of experimental teaching more accurately and intelligently. The algorithm model can improve the overall teaching level.

5. Conclusion

Digital Intelligent ERP experimental teaching is a new teaching mode. It is currently in first operation, and there are still many areas need to be improved. In this study, we combine SVM algorithm with Adaboost algorithm to better strengthen the experimental teaching effect of immersive economic and management courses. At the same time, it intelligently predicts more suitable teaching methods for future newly added course content. Our studies used relatively small sample size, and the research process is relatively uncomplicated. However, it is still a method that provides reference significance for future immersive experimental teaching of digital intelligent processing and management.

The experimental index in this study is to verify the effect of ERP application of digital intelligent SVM-Adaboost algorithm. In fact, a large number of SVM-Adaboost algorithm is used to strengthen teaching prediction. In order to truly improve the accuracy of prediction, a certain amount of other indexes need to be increased. Only with more sufficient indexes and sample data can higher quality SVM-Adaboost training model be obtained better, and eventually improve the prediction and feedback accuracy.

References

[1] Couillet, R.; Liao, Z. Random Matrix Methods for Machine Learning; Cambridge University Press: Cambridge, 2022.

[2] Jiang, H. Machine Learning Fundamentals: A Concise Introduction; Cambridge University Press: Cambridge, 2021.

[3] Wu, J. Classifiers and Tools. In Essentials of Pattern Recognition: An Accessible Approach (pp. 143-172). Cambridge: Cambridge University Press. 2020.

[4] Joseph, A. D.; Nelson, B.; Rubinstein, B. I. P.; Tygar, J. D. In Adversarial Machine Learning; Cambridge University Press: Cambridge, 2019; pp 167–198.

[5] Bhandari, GP, Gupta, R., and Upadhyay, S. K. "A Method for Fault Prediction in SOA Based Systems Using Machine Learning Technology, Data Technology and Applications, Vol. 53, No. 4, pp. 397-421, 2019.

[6] Souza, F. ; Couto, B. ; Silva, G. H. S. da; Dias, I. G. ; Rigueira, R. V. M. ; Pimenta, G. M. ; Martins, M. ; Mendes, J. C. ; Braga, G. M. ; Teixeira, J. A. ; Santos, R. C. ; Martins, J. M. C. ; Sousa, K. S. de; Souza, D. N. de; Alves, G. B. ; Rocha, V. A. R. Prediction of Surgical Risk in General Surgeries: Process Optimization Through Support Vector Machine (SVM) Algorithm. Infection Control & Hospital Epidemiology 2020, 41 (S1), s355–s356.

[7] Vazquez, J.; Abdelrahman, S.; Byrne, L. M.; Russell, M.; Harris, P.; Facelli, J. C. Using Supervised Machine Learning Classifiers to Estimate Likelihood of Participating in Clinical Trials of a De-Identified Version of Research Match. Journal of Clinical and Translational Science 2021, 5 (1), e42.

[8] Zhang, A., Zhang, P. and Feng, Y. "Short term load forecasting of microgrids based on DA-SVM", Compulsory - International Journal of Electrical and Electronic Engineering Computing and Mathematics, Vol. 38 No. 1, pp. 68-80, 2019.

[9] SHAKERIN, F. A. R. H. A. D.; GUPTA, G. O. P. A. L. White-Box Induction From SVM Models: Explainable AI with Logic Programming. Theory and Practice of Logic Programming 2020, 20 (5), 656–670.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230193

Research on the Influence of Intelligent Manufacturing on R&D Efficiency of Manufacturing Enterprises–Based on the Quasi-Natural Experiment of China's Intelligent Manufacturing Demonstration Project

Zichen Guo and Dan Li[1]
School of Economics and Management, Xiangnan University, Chenzhou, China

Abstract. Using the quasi-natural event of the promotion of China's intelligent manufacturing demonstration project, this paper takes China's A-share listed manufacturing enterprises from 2011 to 2021 as the research sample, takes R&D efficiency as the explained variable, and dummy variable during the experimental period as the explanatory variable. With enterprise age, fixed capital density, cash flow, enterprise growth, asset-liability ratio, return on equity and enterprise size as control variables, propensity score matching and difference-in-differences model (PSM+DID) were used to empirically test the influence of intelligent manufacturing on R&D efficiency of manufacturing enterprises. The research results show that intelligent manufacturing can effectively improve the R&D efficiency of enterprises in pilot demonstration projects, forming a demonstration and promotion effect, and these benchmarking enterprises promote the high-quality development of China's manufacturing industry. The empirical results enrich the research content of intelligent manufacturing, provide micro-evidence for China's policy formulation and effect evaluation, and also put forward countermeasures and suggestions for improving the R&D efficiency of manufacturing enterprises, so as to promote the process of "intelligent manufacturing" in China.

Keywords. Intelligent Manufacturing, Manufacturing Enterprises, R&D Efficiency

1.Introduction

The manufacturing industry is the main body of the national economy, the foundation of building the country, the tool of rejuvenating the country and the foundation of strengthening the country. With the accelerated development of modern information

[1] Corresponding Author, Dan Li, School of Economics and Management, Xiangnan University, Chenzhou, China; E-mail:187237129@qq.com.

It is the research result of The Key Cultivation Base for "The 14th Five-Year Plan" of Educational and Scientific Research (Lifelong Education Research Base(Fundamental Theory Area))(grant no.XJK22ZDJD58) in Hunan Province.

technologies such as artificial intelligence and big data, China's "big but not strong" manufacturing industry urgently needs to transform and upgrade to "intelligent manufacturing". In 2015, the Chinese government issued "Made in China 2025", proposing to accelerate the intelligent transformation of production equipment in the manufacturing industry. In 2021, eight departments including the Ministry of Industry and Information Technology and the National Development and Reform Commission, issued the "14th Five-Year Plan for Intelligent Manufacturing Development" (hereinafter referred to as "Intelligent Manufacturing Development Plan"), putting forward the "two-step" strategy for China's intelligent manufacturing. This series of important deployment files indicates that intelligent manufacturing will lead a new round of manufacturing revolution. China's manufacturing industry will focus on intelligent manufacturing, with an expectation to improve product quality and seize the commanding heights of future economic and technological development by superimposing intelligent upgrading of the manufacturing industry, so as to become a "big and strong" Chinese manufacturing industry.

At the 20th National Congress of the Communist Party of China, it was pointed out that China has become an innovative country by implementing the new development philosophy and promoting high-quality development. By 2021, China's R&D investment intensity ranked second in the world, second only to the US. With sustained policy support, China has cultivated a number of outstanding enterprises with core competitiveness. However, research and development investment is an investment project with high risk, high failure rate and difficulty in generating the ROI in the short term (Xie Weimin, 2017)[1]. There are not many companies that are truly committed to research and development and are not afraid to fail. Therefore, R&D efficiency not only needs to be guaranteed and improved by internal control (Wang Yunchen et al., 2015), fund management (Hu Yuanmu, 2012), personnel systems (Liang Laixin and Ma Rufei, 2009) and other management systems, but also needs a series of government policies to truly encourage enterprises to invest in R&D (Akethof, 2009; Lu Guoqing et al., 2014; He Qing et al., 2022)[2][3][4][5][6][7]. Scholars have found that different types of financial science and technology subsidies (Ma Jiannan et al., 2018) and government R&D subsidies (Lai Fenghui et al., 2021) have produced positive effects on R&D input [8][9].

Intelligent manufacturing has become an emerging production mode in the manufacturing industry. Through the 14th Five-Year Plan for Intelligent Manufacturing Development, the Chinese government selects a number of pilot demonstration projects of intelligent manufacturing in the first stage, and reduces the product development cycle before and after implementation, reduces operating and production costs, builds intelligent benchmarking enterprises, summarizes and forms effective experience and models, and forms demonstration and promotion effects. Based on the quasi-natural event of the implementation of intelligent manufacturing, scholars have conducted empirical tests on the aspects of enterprise performance (Ying Limeng et al., 2020; Chen Jinliang et al., 2021), cost stickiness (Quan Xiaofeng, 2022) and average wage of manufacturing industry (Yang Xiaofeng, 2018) conducted empirical tests and found that the promotion of intelligent manufacturing is conducive to improving enterprise performance and average wage and reducing cost stickiness [10][11][12]. However, in the field of intelligent manufacturing research, few literatures have tested the effect of this policy. Therefore, this paper uses the quasi-natural experiment of China's intelligent manufacturing demonstration project, takes China's A-share manufacturing listed enterprises as research samples, and adopts propensity score matching and difference-in-differences model (PSM+DID) to empirically study and explore the effect of intelligent manufacturing on enterprise R&D efficiency. We look forward to enriching the research content of intelligent manufacturing, and provide reference for the formulation of follow-up promotion policies and effect evaluation of China's intelligent manufacturing development plan.

2. Research Design

2.1 Research Hypothesis

In the past decade, the Chinese government has consistently stressed economic development on the real economy and encouraged manufacturing enterprises to increase investment in research and development through fiscal, tax and financial means. In accordance with the requirements of the development plan, the Chinese government has strengthened overall coordination between the central and local governments to provide guarantees for pilot enterprises from three aspects: finance, public services, and opening-up and cooperation. In accordance with the requirements of the development plan, the Chinese government has strengthened overall coordination between the central and local governments to provide guarantees for pilot enterprises from three aspects: finance, public services, and opening-up and cooperation. It is worth exploring whether the R & D efficiency of enjoying financial and public services will be significantly improved due to this policy. Therefore, by examining the relationship between intelligent manufacturing and R&D efficiency, this paper further examines whether the policy has achieved substantial improvement in R&D efficiency. This paper puts forward the hypothesis:

H1: Compared with enterprises not included in the intelligent manufacturing pilot, the R&D efficiency of intelligent manufacturing pilot enterprises has been significantly improved.

Similarly, the Chinese government has implemented a series of fiscal and tax policies to encourage R&D investment of manufacturing enterprises. Whether the R&D efficiency of policy pilot enterprises is significantly better than that of other non-pilot enterprises is also examined in this paper. Therefore, the hypothesis is proposed:

H2: There is no significant difference between the R&D efficiency of the intelligent manufacturing pilot enterprises and those not included in the intelligent manufacturing pilot enterprises.

2.2 Sample Selection and Data Sources

This paper takes China's A-share manufacturing listed enterprises from 2011 to 2021 as research samples. According to the intelligent manufacturing pilot demonstration projects published by the Ministry of Industry and Information Technology over the years, the listed companies implementing the intelligent manufacturing pilot projects are identified. In this paper, Stata software is used to process the sample as follows: (1) Samples with missing data in variables were eliminated; (2) The tail of the continuous variable is reduced at the 1% and 99% levels. After the above treatment, 13,058 "enterprise-year" samples were preliminarily obtained, including 87 experimental enterprises (treat) that implemented intelligent manufacturing during the sample period, and 2330 control enterprises that did not implement intelligent manufacturing during the sample period. The above sample enterprise financial data are from the GUOTAI JUNAN database.

a) Explained Variable

Based on Accounting Standards for Business Enterprises and existing studies (Zhai Shuping, Gu Qun, 2013), this paper defines R&D investment (rd) as the explained variable and represents the ratio of R&D investment to total assets at the beginning [13].

b)Explanatory Variable

In this paper, treat is defined as a dummy variable in the experimental period, indicating whether or not the sample company belongs to the experimental group. In this paper, according to the intelligent manufacturing pilot demonstration project publicized by the Ministry of Industry and Information Technology, the implementation of intelligent manufacturing pilot listed companies identified as the experimental group, the value is 1, otherwise belong to the control group, the value is 0. The dummy variable post represents the time before and after the implementation of the intelligent manufacturing project of the sample company. The value before implementation is 0 and after implementation is 1. The cross-multiplying term treat×post was defined as a dummy variable to measure the R&D efficiency of the experimental group.

c) Control Variable

controls are the control variables. According to the existing literature, this paper controls age (age), fixed capital density (fcd), cash flow (cashflow), enterprise growth (growth), asset-liability ratio (debt), return on equity (roe) and enterprise size (size) respectively. Detailed definitions of specific variables are shown in Table 1.

Table 1. Variable Definition

	Symbol	Variable	Variable definition
Explained Variable	rd	R&D investment	R&D investment/total asset in the beginning year
Explanatory Variable	treat	Experimental group/control group	The listed companies implementing the intelligent manufacturing pilot are determined to be the experimental group, with the value of 1; otherwise, they belong to the control group, with the value of 0.
	post	Whether the enterprise implements intelligent manufacturing dummy variables	The production mode of intelligent manufacturing is disclosed for the first time in the annual approval disclosure of the enterprise, and the value is 1 in the current year and subsequent years, otherwise the value is 0.
Control Variable	age	Age	From company establishment to sample time
	fcd	Fixed capital density	Total asset in the ending year/sales revenue
	Cash flow	Cash flow	Net operating income/total asset in the beginning year
	growth	Enterprise growth	Tobin's Q
	debt	asset-liability ratio	Total debt/Total asset
	roe	Return on equity	Net income/Net equity
	size	Size	The natural log of the firm's total assets in the ending year

2.3 Model Design

In order to test the relationship between the implementation of intelligent manufacturing and R&D efficiency in H1 enterprises, a difference-in-differences model is constructed by referring to existing practices:

$$\mathrm{rd}_{it} = \beta_0 + \beta_1 \mathrm{treat}_{it} \times \mathrm{post}_{it} + \beta_j \sum_{j=1}^{n} controls + u_i + \lambda_t + \varepsilon_{it}, \quad i = 1,2,...,n; t = 1,2,...,T \quad (1)$$

Model (1) is a multi-period difference-in-differences model controlling individual and time points. Among them, the explained variable rd is the research input; controls is the sample control variable; u_i, λ_t are respectively individual and time-point fixed effects.

3. Empirical Test and Analysis

3.1 Propensity Score Matching (PSM)

In order to control the selectivity bias of samples, this paper adopts the propensity matching method (PSM) to pair the basic samples. In this paper, based on policy documents such as *Intelligent Manufacturing Development Plan (2016-2020)*, matching variables including enterprise age (age),size (size), return on equity (roe), growth (growth),and asset-liability ratio (debt) are selected as matching variables.

According to the 1:1 nearest neighbor matching principle, we found similar control groups for each experimental group, and finally got 626 paired samples (1252). We further use the annual report, the company's official website and internet search for manual inspection of the control group enterprise, if the content of public reports has nothing with intelligent manufacturing, we define it as the control group enterprise which has not implemented intelligent manufacturing, to ensure that the control group enterprise has not implemented intelligent manufacturing.

Finally, we conducted a balance test on the paired samples, and the results are shown in Table 2. As can be seen from Table 2, there is no significant difference between the experimental group and the control group in terms of company size, return on equity, company age, company age, etc., which can better alleviate sample self-selection bias, and it is more reasonable to apply matched samples to conduct subsequent DID tests.

Table 2. PSM balance test

variables	treat	control	t-value	p-value
age	17.885	14	1.33	0.185
size	22.976	0.00824	4.87	0.000
roe	0.9389	0.06429	0.63	0.531
growth	1.8069	1.9779	-0.38	0.704
debt	0.46904	0.25088	2.39	0.017

3.2 Descriptive Statistics

The descriptive statistical results of main variables are shown in Table 3. The average R&D investment (rd) of the sample companies was 0.032, i.e. the average R&D investment to total assets ratio of the sample enterprises is 0.032 The results of control variables show that the average size of companies is 21.62, indicating that the size of companies matched by PSM in this paper is large, which is consistent with the characteristics of the manufacturing industry; The average asset-liability ratio (debt) is 0.356, indicating that the debt of the sample companies is at a reasonable level. Besides, the distribution of other variables is within a reasonable range.

Table 3. Descriptive statistics

variables	N	mean	sd	min	max
rd	1,252	0.0320	0.0252	0	0.367
age	1,252	17.58	5.721	5	33
fcd	1,252	2.070	1.185	0.601	8.238
cashflow	1,252	0.0557	0.0667	-0.136	0.252
growth	1,252	2.197	1.206	0.957	7.101
debt	1,252	0.356	0.206	0.0470	0.864
roe	1,252	0.0768	0.123	-0.548	0.360
size	1,252	21.62	1.643	19.35	25.35

3.3 Basic Regression Analysis

Table 4 shows the results of differential regression. Column (1) is the regression of the paired samples after the adoption of propensity score matching (PSM). The estimated coefficient of treat×post for the cross-multiplying term is 0.009, which is significant at the 1% level, indicating that the R&D efficiency of intelligent manufacturing pilot enterprises has been improved more significantly than that of 626 control enterprises with no significant difference. Column (2) is the regression using the full sample, and the estimated coefficient of treat×post for the cross-crossing term is 0.008, significant at the 1% level, indicating that the R & D efficiency of intelligent manufacturing pilot enterprises has been significantly improved compared with other manufacturing enterprises in the full sample of manufacturing enterprises.

With a series of supporting policies, Chinese manufacturing enterprises have significantly increased their willingness to invest in research and development. For enterprises in the intelligent manufacturing development plan, the scale and proportion of R&D investment are significantly higher than other manufacturing enterprises. The regression results support hypothesis H1. Compared with enterprises not included in the intelligent manufacturing pilot, the R&D efficiency of intelligent manufacturing pilot enterprises is significantly improved.

The control variables show that the fixed capital density (fcd) of manufacturing enterprises is significantly negatively correlated with the R&D efficiency at a 1% level, that is, the higher the ratio of heavy assets of the sample companies, the lower the R&D efficiency of the enterprises. There is a significant negative correlation between the size of the company and the R&D efficiency. The results show that the larger the size of the company, the R&D efficiency will decrease. The growth of manufacturing enterprises is significantly positively correlated with R&D efficiency, that is, the larger Tobin's Q is, the higher the R&D efficiency is.

Table 4. Regression results of intelligent manufacturing and R&D efficiency

	(1)	(2)
variables	**rd**	**rd**
treat×post	0.009***	0.008***
	(5.42)	(6.13)
age	-0.000	-0.001***
	(-0.84)	(-5.37)
fcd	-0.003***	-0.003***
	(-4.19)	(-7.75)
cashflow	0.016	-0.006
	(1.43)	(-1.58)
growth	0.002***	0.003***
	(2.88)	(9.11)
debt	0.004	0.001
	(0.95)	(0.15)
roe	0.004	-0.005
	(0.59)	(-0.58)
size	-0.003***	-0.002***
	(-4.66)	(-2.88)
Constant	0.086***	0.070***
	(7.61)	(5.03)
Observations	1,252	13,055
R-squared	0.076	0.051

Note: 1. ***, **, * are represent significance at 1%, 5% and 10% levels respectively; 2. the t values are in parentheses.

3.4 Placebo Test

In order to further verify the hypothesis, a placebo test was conducted for the re-selection of experimental time. If the improvement of R&D efficiency is caused by the effect of the policy implementation of the "14th Five-Year Plan for Intelligent Manufacturing Development", then the estimated results of the difference-in-differences model will no longer be valid if the policy time is changed. Therefore, we refer to the practice of Quan Xiaofeng and Li Chuang (2022), by which the introduction of intelligent manufacturing was brought forward by two years to re-observe the relationship between intelligent manufacturing and R&D efficiency. The results of the placebo test in Table 5 showed that the regression coefficient of treat×post in column (1) was no longer statistically significant at 0.001, indicating that the improvement of R&D efficiency of enterprises in the experimental group was caused by the effect of the implementation of intelligent manufacturing development planning.

3.5 Robustness Test

In order to ensure the robustness of the research results, we used alternative R & D investment measures to test. Based on the idea of Aschhoff (2009), R&D investment intensity (rdint) --R&D expenditure/ operating income, was selected. According to the regression results of the robustness test in column (3) of Table 5, the regression coefficient of treat×post, the cross-multiplying term of rdint, is 0.018 and significant at the 1% level, and the regression results support hypothesis H1. Column (4) is the robustness test of the whole sample. The treat×post estimation coefficient of the cross-multiplication term is 0.017, which is significant at the 1% level. The regression results still support hypothesis H1.

We also conducted a placebo test on the substitute index, and the regression results were shown in column (2) in Table 5. The regression coefficient of treat×post was still not significant. The improvement of R&D efficiency of enterprises in the experimental group was caused by the effect of the implementation of intelligent manufacturing development planning.

Table 5. Placebo Test and Robustness Test

	placebo test	**placebo test**	**robustness test**	**robustness test**
	(1)	**(2)**	**(3)**	**(4)**
variables	**rd**	**rdint**	**rdint**	**rdint**
treat×post	0.001	0.003	0.018***	0.017***
	(-0.46)	(-0.72)	(-5.46)	(-8.29)
age	-0.000***	-0.001***	-0.000*	-0.000***
	(-3.40)	(-4.02)	(-1.75)	(-7.89)
fcd	-0.003***	0.014***	0.015***	0.013***
	(-5.02)	(-7.83)	(-7.91)	(-24.93)
cash flow	0.007	-0.018	0.003	-0.005
	(-0.62)	(-0.76)	(-0.11)	(-0.87)
growth	0.002***	0.005***	0.005***	0.005***
	(-2.81)	(-4.05)	(-4.08)	(-12.58)
debt	0.001	-0.018**	-0.011	-0.025***
	(-0.2)	(-2.10)	(-1.27)	(-10.45)
roe	0.005	-0.009	-0.013	-0.010**
	(-0.91)	(-0.67)	(-0.96)	(-2.47)
size	-0.001	-0.001	-0.001	-0.002***
	(-1.55)	(-0.62)	(-0.62)	(-4.69)
Constant	0.048***	0.037*	0.037*	0.068***
	(4.14)	(-1.76)	(-1.76)	(-9.22)
Observations	1,252	1,252	1,252	13,055
R-squared	0.117	0.236	0.236	0.216

Note: 1. ***, **, * are represent significance at 1%, 5% and 10% levels respectively; 2. the t values are in parentheses.

In order to study the impact of intelligent manufacturing on R&D efficiency, the ideal method is to compare the R&D efficiency of the same enterprise in the environment of intelligent manufacturing and non-intelligent manufacturing. In reality, however, this environment cannot be realized. Therefore, the above research is a quasi-natural experiment using China's smart manufacturing demonstration project, using the propensity score matching method (PSM) to construct "counterfactuals" to control the selection bias of the sample, and to estimate The influencing effect of R&D efficiency of China's A-Share Manufacturing Listed Companies.

The following results can be obtained by combining Table 4, Table 5 :

(1) Through the empirical test of the paired samples with propensity score matching (PSM), the treat×post estimation coefficient of the cross-multiplying term is 0.009 and significant at the 1% level (Table 4 Column (1)), indicating that the R&D efficiency of intelligent manufacturing pilot enterprises is significantly improved compared with enterprises not included in intelligent manufacturing pilot enterprises, and the regression results support hypothesis H1. This result also passed the placebo test (Table 5 column (1)) and the robustness test (Table 5 column (3)). It shows that after controlling the endogenous problem, the conclusion that the significant improvement of the R&D efficiency of intelligent manufacturing pilot enterprises is caused by the effect of the implementation of intelligent manufacturing development plan is still valid, and the results are stable.

(2) We also conducted the same empirical test on the full sample of China's A-share manufacturing listed enterprises, and the estimated coefficient of cross-multiplying term treat×post was 0.008 and significant (Table 4 column (2)), indicating that in the full sample of manufacturing enterprises, the R&D efficiency of intelligent manufacturing pilot enterprises was significantly improved compared with other manufacturing enterprises. The results were robust by the placebo test (Table 5 column (2)).

In recent years, the Chinese government has introduced a series of policies to support the high-quality development of enterprises to support enterprises in cultivating core competitiveness and accelerate the upgrading of China's industries. Considering the extension of policies in reality. The focus of our attention is whether the "14th Five-Year Plan for Intelligent Manufacturing Development" issued by the Chinese government in 2021 will significantly improve the R&D efficiency of pilot enterprises . The empirical results show that in the context of the improvement of R&D efficiency of Chinese manufacturing enterprises, the R&D efficiency of enterprises included in the development plan of intelligent manufacturing is significantly higher than that of other manufacturing enterprises, playing a role in demonstration and promotion.

3.6 Analysis of Research Results and Suggestions of Solution

The 14th Five-Year Plan for Intelligent Manufacturing Development adopts the pilot plan before the promotion to improve the intelligent development of China's manufacturing industry. The research results show that the promotion of intelligent manufacturing demonstration projects in China has achieved a good demonstration effect. Compared with other manufacturing enterprises, the R&D efficiency of enterprises included in the intelligent manufacturing pilot has been significantly improved. It can be seen from the side that in the past seven years, China's central government and local governments have provided guarantees for pilot enterprises from the aspects of finance, public services, and opening-up and cooperation according to the requirements of development planning, and achieved good results. The research and development efficiency of pilot enterprises has been significantly improved due to this policy.

This paper put forward two suggestions based on the above research results:

(1) From the government's point of view, China should firmly focus on the real economy. We will support the intelligent development of manufacturing enterprises through more comprehensive and fair safeguard measures in terms of improving the innovation system, fiscal and tax support, and innovative finance. Some works need to be done to reduce the risk of enterprise innovation and research in many ways, and improve the willingness of enterprises to research and development so that enterprises

committed to research and development can achieve rapid development with the support of the government.

(2) From the perspective of enterprises, under the support of a series of government policies, enterprises should improve the level of internal governance, accurately grasp the direction of industrial development, and effectively improve the treatment and environment of research and development personnel, so that the research and development results could generate new revenue and become the moat of enterprises.

4. Conclusion

This paper uses stata software to make a difference-in-differences empirical analysis of the R&D efficiency of manufacturing enterprises after addressing the endogeneity problem by propensity score matching method. The research results show that in recent years, manufacturing enterprises have invested more and more in R&D. Under the implementation of an intelligent manufacturing development planning policy, the R&D efficiency of pilot enterprises is significantly better than that of other manufacturing enterprises, and the prominent demonstration and promotion effect is achieved. This paper studies the pilot enterprises as a benchmark to promote the high-quality development of China's manufacturing industry. The conclusion of this study enriches the research content of intelligent manufacturing and provides micro evidence for the formulation and effect evaluation of the follow-up promotion policies of China's intelligent manufacturing development plan. This paper selects the quasi-natural experiment of an intelligent manufacturing demonstration project to study the impact of intelligent manufacturing on the R&D efficiency of manufacturing enterprises. The scope of the study is relatively limited, and it needs to follow the main line of R&D efficiency investigation and expand the research level in the later period. It is much better to link the implementation of intelligent manufacturing with capital operation management practices such as risk management, investment management, financing management, and supply chain management.

Acknowledgment

This paper is supported by ①The Key Cultivation Base for "The 14th Five-Year Plan" of Educational and Scientific Research (Lifelong Education Research Base(Fundamental Theory Area))(grant no.XJK22ZDJD58)) in Hunan Province.②Hunan Province colleges and universities' teaching reform research project (Hunan Provincial Department of education [2019] No.291-Item No.867).③the Chenzhou social science planning project No. Czsskl2022103.

References

[1] Xie Weimin. Does the tournament encourage or inhibit corporate innovation[J].China Soft Science,2017,(10): 104 -113.
[2] Wang Yun, Chen Ludong, Gong Yifei. Does internal control improve R&D efficiency?[J]. Securities Market Herald,2015,(10): 39-45.

[3] Hu Yuanmu. Can technical independent Directors Improve R&D output Efficiency?—From the Chinese securities market research[J].Nankai Business Review,2012,(2):13-142.
[4] Liang Laixin, Ma Rufei. R&D Capital management and enterprise independent innovation-based on the empirical analysis of information technology listed companies[J]. Journal of Finance and Economics,2009,(B):49-59
[5]Akethof，B.2009 “The Effect of Subsidies on R&D Investment and Success-lo Subsidy History and Size Matter?” [J].ZEW DiscussionPapers，No.09-032.
[6] Lu Guoqiang, Wang Zhou, Zhang Chunyu.Research on the performance of Government innovation Subsidies for China's strategic emerging Industries [J].Economic Research Journal,2014(7):44-55.
[7] He qing, Liu Jingran, Fan Qingquan.Research on enterprise R&D risk and subsidy policy optimization[J].Economic Research Journal,2022,(5): 192-208.
[8] Ma Jianan, Zhai Haiyan, Dong Jing. An empirical study on the influence of financial science and technology subsidies on enterprise R&D input[J].Public Finance Research,2018(2):77-87.
[9] Lai Fenghui, Li Shanmin, Wang Dazhong.Design of government R&D subsidy mechanism under the constraint of enterprise financing[J].Economic Research Journal,2021(11):48-66.
[10] Chen Jinliang, Zhao Yaxin, Lin Song.Can intelligent Manufacturing promote enterprise innovation Performance? [J].Foreign Economics&Management,2021(9):83-101.
[11] Quan Xiaofeng, Li Chuang.Intelligent Manufacturing and cost stickiness -- a quasi-natural experiment from China's Intelligent Manufacturing Demonstration Project[J].Economic Research Journal,2013,(2): 68-84.
[12] Yang Xiaofeng.Will intelligent manufacturing help raise average manufacturing wages?--based on industrial robot data of 17 provinces from 2001 to 2016[J].Economic Research Journal,2022,(5): 169-176.
[13] Zhai Shuping, Gu Qun. Research on financial development, financing constraint easing and R&D investment efficiency of high-tech enterprises[J].Jingji Jingwei Economic Survey,2022,(4): 138-143.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230194

Artificial Intelligence-Based Hybrid Recommendation Algorithm for Learning Resources

Weibo Huang[a, 1], Mohan Liu[b] and Hao Chen[c]
[a] *Modern Education Technology Center, Experimental Teaching Center, Guangdong University of Foreign Studies, Guangzhou, China.*
[b] *School of English Education, Guangdong University of Foreign Studies, Guangzhou, China.*
[c] *School of Information Science and Technology, School of Cyberspace Security, Guangdong University of Foreign Studies, Guangzhou, China.*

Abstract. To meet the development needs of Chinese learners' personalized learning in the context of "Internet+" educational applications, this paper proposes to fuse the topic model (LDA)with the convolutional neural network mode (CNN). It helps us obtain the topic information and semantic information of different learning resources. Then we can obtain the corresponding learning resource feature vector. Secondly, we use the TF-IDF and K-nearest neighbor algorithms to vectorize the different persona information and potential interests of the learners. Then we introduce above-mentioned learning resource feature vectors, use the cosine similarity of the multidimensional vectors to calculate the learning interaction degree between the learners and the learning resources. Finally, we use a collaborative filtering recommendation algorithm based on the K-nearest neighbor algorithm to combine the knowledge point feature vectors of learning resources and learners' results. This study can realize personalized recommendation of learning resources for learners.

Keywords. Personalized learning, Thematic model, Convolutional neural network model, TF-IDF algorithm

1. Introduction

Based on Artificial Intelligence technology and education theory knowledge, the companion drama-oriented intelligent teaching system combines Foreign language learning with the development of the companion drama-oriented system. It changes the traditional Foreign language learning method with unique companion drama-oriented features. Besides, it eliminates the obstacles and thought burdens caused by artificial emotional factors in the process of natural conversation practice of general learners. It not only improves learners' overall Foreign language proficiency and learning efficiency,

[1] Corresponding Author: Weibo Huang, Modern Education Technology Center, Experimental Teaching Center, Guangdong University of Foreign Studies, China; This study is one of the research achievements of the Guangdong Higher Education Teaching Research and Reform Project in 2021, and is also financially supported by the Undergraduate Innovation Training Program Project of Guangdong University of Foreign Studies in 2023. Email: hwb444@163.com

but also makes Foreign language learning more interesting and helps to increase learners' initiative in learning Foreign language. In addition, it helps learners change the "search-and-learn" learning style and develop their collaboration, communication (questioning) and other skills through "dialogic learning", and develop their critical and creative thinking.

2. Current status of domestic and international research

Some researchers have tried to use Metaverse to build language teaching scenarios [1], such as by building a Metaverse simulation language learning space to promote teaching reform [2]. Other researchers have used Metaverse for interactive collaboration in online classes [3] to explore its pedagogical effects. Some educational institutions have currently collaborated with Second Life platforms to build learning environments [4], such as some researchers focusing on the design and implementation of immersive learning environments [5]. The emergence of Metaverse meetings overcomes a series of problems that existed during previous virtual space interactions, such as too little interaction, screens blocking emotional communication [6], and poor learning outcomes [7]. It allows multiple users to use virtual reality technology to enter the virtual meeting room for interaction [8]. This gives users an immersive experience and optimizes the overall effect of the meeting. In addition, it has great potential as a learning and practice platform for communication and dialogue [9]. It provides an economical, individualized dialogue scenario [10]. In this context, learners can overcome their lack of confidence in pronunciation, grammatical correctness, and the shyness that most of them may have. This helps them enter an effective "conversation". It is particularly positive in terms of taking the first steps towards "dialogue".

3. Current status of domestic and international research

The model of the hybrid recommendation system consists of three main components and a set-reference impact factor β: existed information model, potential interest model, hybrid recommendation algorithm model, and information about the characteristics of the learning group corresponding to the learning resources (impact factor β). Then, we define the above three components and set the reference impact factors here.

Existed information model (EIM): In deep learning, diverse raw data inputs can cause the distribution of feature values to vary widely. It is necessary to convert it into a form that can be recognized by the framework in advance. This operation, to some extent, can effectively respond to the basic personal information of different learner's initialized multiple educational backgrounds, learning abilities, self-perceptions, emotional tendencies, and interest levels. They are. We quantified age, gender, education, profession, identity, style, emotion, and hobby as feature vectors. And after the subsequently described TF-IDF algorithm for calculating the feature weights, we finally obtained the weight vectors of their different feature values. They are denoted here as EIM = {e11, e12, e13, ..., e18}, and called the existed information model.

Potential interest model (PIM): The PIM is defined here to identify the neighbors of the current user by using the KNN collaborative filtering algorithm as described later. Neighbors are groups of users who have similar interests to the current learner and have common learning resources to utilize. We define the potential interest model as the existed information model of its neighbors who have similar interests and are members of the same learning user group. And after the subsequently described TF-IDF algorithm for calculating the feature weights, we finally obtained the weight vectors of their different feature values. They are denoted as PIM = {p11, p12, p13, ..., p18}, and is called the potential interest model.

Hybrid recommendation algorithm model (HRAM): Hybrid recommendation algorithm model is the weight vector. We merge the above two models (EIM+PIM) according to certain logical rules and mathematical operations to get it (in this paper, we use the calculation of cosine similarity). They are denoted as H1 = {h11, h12, h13, ..., h18}, and is called the hybrid recommendation algorithm model. Finally, it combines with the consideration of the factors affecting the numerator and presents the personalized recommendation reference content generated by a hybrid recommendation system.

For the hybrid recommendation algorithm model, this paper presents here an impact factor β that can improve its effectiveness: a model of the characteristic information of the learning group corresponding to the learning resource. Based on the characteristic information of the learning group corresponding to the learning resource, it is possible to understand the type of learning group corresponding to that type of learning resource. This can effectively reflect the general characteristics of the users learning the resource, so that the hybrid algorithm can make constant tuning recommendations. The evaluation formula of impact factor β:

$$\beta = \frac{\overline{M1}}{sim(x,y)} \tag{1}$$

$\overline{M1}$ denotes the average of the multidimensional weight vector of the base information that the learner itself already has for measuring itself (which is not the same for each learner). $sim(x, y)$ denotes the similarity between the two models, where x and y can be M1 or M2 (but not both) respectively, and its calculation method in this study takes the calculation of cosine similarity. That means $sim(x, y)$ becomes larger, x and y are more and more similar with each other. Therefore, the role of the impact factor β explored in this paper refers to the fact that the hybrid recommendation algorithm model merges the above two models (EIM+PIM) in the subsequent according to certain logical rules and mathematical operations. If the impact factor β is smaller at this point, it indicates that the content of the generated recommendation results is more accurate. Then it is possible to make recommendations to user learners. On the contrary, it indicates that the content of the generated recommendation results is more deviant and does not conform to the general characteristics of the learning groups corresponding to such learning resources. Finally, it will present a more accurate personalized recommendation reference content generated by a hybrid recommendation system.

The overall framework of the hybrid recommendation system combines a content-based recommendation algorithm and a KNN-based collaborative filtering algorithm. It uses the KNN algorithm in the beginning data preprocessing part and the final part of the recommended learning resources. After that it uses an LDA topic model interacting with a deep convolutional neural network CNN model for learning to analyze and classify the proposed recommended learning resources in an enhanced way. Finally, it combines the characteristic information models (impact factors) of the learning groups corresponding to the learning resources to recommend suitable learning resources to different learners.

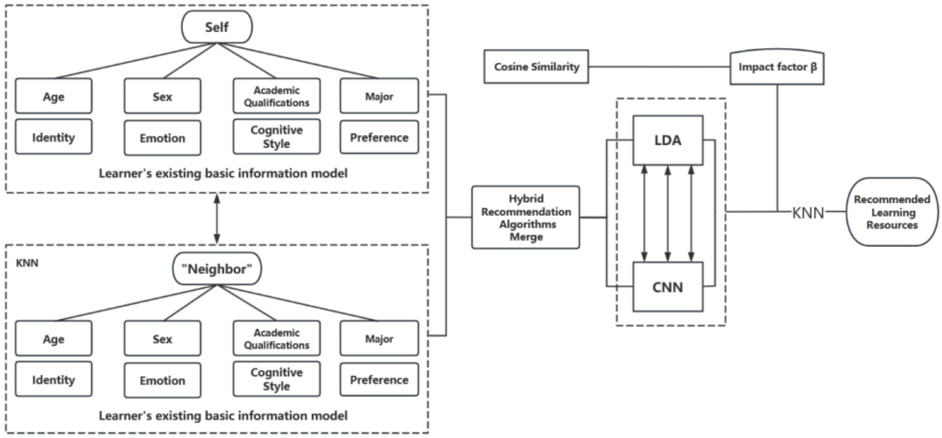


Figure 1. The overall framework of the hybrid recommendation system.

In building the above existed information model and potential interest model, the TF-IDF (Term Frequency Inverse Document Frequency representation) algorithm vectorization is needed to preprocess the original input data of the learner and calculate its feature weights. The purpose of data preprocessing is to optimize the model for the raw data at the input side to be more suitable for subsequent neural network processing.

$$W_{X,Y} = tf_{X,Y} \times \log\left(\frac{N}{df_x}\right) \tag{2}$$

In TF-IDF, for each indicator, $tf_{X,Y}$ denotes the frequency of occurrence of this indicator among all Foreign language learners. N represents the total number of the search content set. And df_x contains the number of samples in which the indicator has appeared. $tf_{X,Y} * \log\frac{N}{df_x}$ is recorded as the result to the feature vector weights of each learner.

The reinforcement classification learning model used in this study, implicit Dirichlet distribution (hereafter abbreviated as LDA), is a topic model. It generates topics based on the word frequencies in a set of documents. It is essentially a three-layer Bayesian structure. They are document layer, topic layer and word layer. Usually, a topic model generally extracts several sets of keywords from a set of documents to express the core idea of its object, i.e., the "topic". LDA is a very effective method for finding an accurate and reasonable mixture of topics in a set of given documents. The neural network-based topic model used in this study mainly uses deep learning to reconstruct the text production process of the topic model. And it adds topic-word sparsity constraints to the modeling process to generate more expressive topic words.

The LDA algorithm adopted in this study is to first normalize the d-dimensional given learning resource dataset (d is the number of features), calculate the d-dimensional mean vector for each class. Then we calculate the eigenvalues of the matrix and the corresponding eigenvectors by constructing the between class scatter matrix S_B and the within class scatter matrix S_W. Finally, the eigenvectors corresponding to this learning material will be subsequently used in the calculation of cosine similarity.

$$S_B = \begin{bmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{n1} & \cdots & p_{nn} \end{bmatrix}$$

(between class scatter matrix S_B)

$$S_W = \begin{bmatrix} e_{11} & \cdots & e_{1n} \\ \vdots & \ddots & \vdots \\ e_{n1} & \cdots & e_{nn} \end{bmatrix}$$

(within class scatter matrix S_W)

The convolutional layer of a convolutional neural network CNN mainly consists of two basic operations, one of it is feature extraction and the other is feature mapping. In this study, we adopt the first feature extraction operation mentioned above. The aim is processing the relevant learning resources in the input layer. Finally, we combine the processed data with the LDA model to derive the corresponding knowledge point feature vectors.

In this study, we specify the input size of the convolutional neural network CNN as 1. It corresponds to the channel dimension of the integer sequence input. Then the input is embedded using a feature vector embedding of dimension 8. For n element vectors to be spaced, layer blocks are created containing convolutional layers, batch normalization layers, ReLU layers, discard layers, and maximum pooling layers. For each block, we specify 100 convolution filters of size 1 × N and a global maximum pooling layer. The input layers connect to each block and the concatenation layer concatenate the outputs of each block. Finally, we classify the output. It includes a fully connected layer with output size K, a Softmax parser layer, and a classification layer. And K is the number of classes.

The $sim(x,y)$ mentioned in the above analysis, indicates the similarity between two models. For the calculation of $sim(x,y)$, the main methods are cosine similarity, Euclidean Algorithm and Pearson correlation coefficient. In this study, we adopt the method of calculating cosine similarity. Based on the collected user data, we use the TF-IDF algorithm to vectorize the original input data of the learners for preprocessing and calculating as the input of IM_X and PIM_Y (or vectors a and b) for calculating the cosine similarity of the underlying information between learners.

$$sim(x,y) = \cos(x,y) = \frac{EIM_X \cdot PIM_Y}{|EIM_X| \cdot |PIM_Y|} \tag{3}$$

Then, we use LDA and CNN fusion analysis to derive the feature vectors of the corresponding classified learning resources. And before, we used the hybrid recommendation algorithm model to merge the two models (EIM+PIM) according to certain logical rules and mathematical operations to obtain the learner vector. Then we use the cosine similarity again to calculate and define the threshold value of similarity γ to find the most similar learning resource recommendation set for each user. Finally, we use the impact factor β to check and verify the accuracy of the recommended learning resources.

At the end of the AI-based hybrid recommendation algorithm for Foreign language learning resources discussed in this paper is a KNN-based collaborative filtering algorithm. It has the same function as the data preprocessing part. Firstly, this paper's algorithm calculates and tests the Euclidean distance “dist” from sample points to all sample points, and then uses the hook theorem to calculate it. We set the parameter K and select the K points closest to the point to be measured. Then we count the number of each type or class label from these K points and select the class label number that appears most frequently as the class label number of the unknown sample. It can feedback the prediction results to recommend the learning resources that are most similar with the users' interactions with knowledge points to the recommended subjects. Finally, it generates a list of academic resource recommendations for each user.

4. Experimental results

In this study, the data of learners were obtained from different types of learners selected by Kaggle for optimal matching. Each learner sample has different information/behavioral characteristics about themselves. The Item ID of the learning resource represents the unique flag serial number of each learning resource. The priority with 1 is the highest, 2 being the second highest and 3 being the lowest. The point where 1 is located at the line segment is the Item ID flag sequence number of the most recommended matching learning resource. According to the analysis of the results, the best learning resource path for a random sample is from learning resource 1 to learning resource 3 and finally to learning resource 2. It finally leads to the recommended personalized learning resource path generated by the hybrid recommendation algorithm system based on the learner's basic personal information.

In this study, we choose the model accuracy (MA) calculation to measure the matching suitability of the recommendation results generated by each algorithmic model. The formula for calculating the model accuracy (MA) can be expressed as follows:

$$MA = \frac{N_a}{N+|N_a-N_{\bar{a}}|} \tag{4}$$

In this formula, N denotes the total number of recommended learning resources in general. N_a denotes the number of recommended learning resources accessed by learner users. $N_{\bar{a}}$ denotes the number of recommended learning resources not accessed by learner users.

This study selected the dataset from the learner information dataset of Kalboard 360. The initial dataset consists of 480 learner records and 16 features. For this study, we optimized this dataset and extracted a sample dataset consisting of 100 learner records and 8 features. For these 100 learners, this study calculated the accuracy of the model by increasing the number of items from 10 to 60. Then we applied the KNN nearest neighbor algorithm to match the neighbor users of size 20, and use the cosine similarity for the calculation of similarity between learning resources and learner users. Finally, this study presented the attempt and innovation of the hybrid recommendation algorithm with the visualized data.

5. Conclusion

For now, it is difficult to combine the functions of situational learning, word frequency analysis, word memorization, and oral communication organically in both theoretical and practical studies. The AI-based companion drama-oriented Foreign language learning system references the multiple characteristics of learners in different dimensions. It visualizes knowledge and provides the necessary feedback and guidance to learners as they engage in thinking, discussion, and other behaviors. This makes learning more enriching and oriented for the learner. It adds interest and practice to learning, slowly increases the difficulty curve of Foreign language learning, and increases learners' enthusiasm for learning. At the same time, this deep learning framework-based learning resource recommendation system is based on the previous resource system, which is continuously updated and iterated on its own. With its polymorphism, inheritability, and expandability, it eventually forms a more accurate personalized resource recommendation to meet the diversified needs of different learners in a targeted manner. It solves the problem of boring and inefficient learners learning Foreign language through the combination of various innovative experiences such as situational experience, AI accompaniment and traditional Foreign language learning methods.

References

[1] Singh R H, Maurya S, Tripathi T, et al. Movie recommendation system using cosine similarity and KNN. International Journal of Engineering and Advanced Technology, 2020;9(5): 556-559.

[2] KLASNJA-MILICEVIC A, IVANOVI M, NANOPOULOS A. Recommender systems in e-learning environments: a survey of the state-of-the-art and possible extensions. Artificial Intelligence Review, 2015;44(4):571-604.

[3] Suzuki, S., Kanematsu, H., & Barry, D. M. et al. Virtual Experiments in Metaverse and Their Applications to Collaborative Projects: The Framework and Its Significance. Procedia Computer Science. 2020;17(6):2125-2132.

[4] GUO Q, MING Z. Implement web learning environment based on data mining. Knowledge-Based Systems, 2009;22(6):439-442.

[5] De Back, T. T., Tinga, A. M., & Louwerse, M. Learning in Immersed Collaborative Virtual Environments: Design and Implementation. Interactive Learning Environments. 2021;1-19.

[6] Raymond Bond, Anne Moorhead, Maurice Mulvenna, Siobhan O'Neill, Courtney Potts, Nuala Murphy. Exploring temporal behavior of App users completing ecological momentary assessments using mental health scales and mood logs. Behavior amp; Information Technology,2019;38(10):1016-1027.

[7] Han, J., & Shi, J. Comparing the Effects of SPOC and Face-to-Face Courses and Factors Influencing the Effects. International Journal of Information and Education Technology. 2022;12(6):565-570.

[8] Vassoyan J, Vie J J, Lemberger P. Towards Scalable Adaptive Learning with Graph Neural Networks and Reinforcement Learning. arXiv preprint arXiv:2305.06398, 2023.

[9] Zhong Q, Yu J, Zhang Z, et al. Towards a General Pre-training Framework for Adaptive Learning in MOOCs. arXiv preprint arXiv:2208.04708, 2022.

[10] Trstenjak B, Mikac S, Donko D. KNN with TF-IDF based framework for text categorization. Procedia Engineering. 2014;69: 1356-1364.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230195

A Study of Current Situation of Knowledge, Attitude, Beliefs and Practices of "Double Reduction" Policy Among Primary and Secondary School Teachers in Chenzhou City

Li YANG, Ling HE[1] and Naizhu HUANG
School of Educational Science and Law, Xiangnan University, 423000, Chenzhou, Hunan Province, China

Abstract. China has been carrying out nationwide efforts to reduce the academic burden of primary and secondary school students over six times from 1949. In June 2021, the Ministry of Education (MOE) established a department to supervise the out-of-school education and training, and suggested to further reduce the homework burden and out-of-school training for students in compulsory education (hereinafter referred to as "Double Reduction"). In order to understand the teachers' literacy in implementing the "double reduction" policy, the study administered questionnaires to 242 primary and secondary school teachers from Chenzhou City, Hunan Province, to survey their knowledge, beliefs, practices and policies. The results showed that primary and secondary school teachers in Chenzhou city have a high degree of awareness, importance, and sustainable implementation of the "double reduction" policy in three aspects viz., students' arrival and departure time, extracurricular training and homework management. Though, the "quantitative change" in the implementation of the "double reduction" policy is obvious, but the "qualitative change" is relatively petty low, and there is a significant difference between urban and rural areas in terms of awareness, recognition and practice. In order to reduce the burden and improve the quality, it is suggested to promote collaboration among schools, government, teachers, parents and other relevant parties to build a new ecology of lifelong education for primary and secondary school teachers, increase teacher training and support for "double reduction", improve the ability of teachers to know, believe and act on the "double reduction" policy and give priority to rural primary and secondary education.

Keywords. primary and secondary school teachers, "double reduction", knowledge, attitude, belief and practice.

[1] Corresponding author, Ling HE, Xiangnan University, China; e-mail: 616728140@qq.com.
This research is the outcome of The Key Cultivation Base for "The 14th Five-Year Plan" of Educational and Scientific Research [Lifelong Education Research Base (Fundamental Theory Area)] in Hunan Province.

1. Problem formulation

In Chinese education system, "double reduction" is short for "effectively reducing homework burden and off-campus training burden of students in compulsory education". Its purpose is to effectively improve the education level of schools and continuously standardize in-school and out-of-school training. Though, China has been reducing the academic burden of primary and secondary school students on six occasions from 1949 but both in-school load and out-of-school load remain unchanged in large number of rural primary and secondary schools. In June 2021, the Ministry of Education (MOE) officially established Department of Supervision of Out-of-School Education and Training which reported in July its Opinions on Further Reducing the Homework Burden and Out-of-School Training for Students in Compulsory Education (hereinafter referred to as "Double Reduction") [1].

Primary and secondary school education subsequently became the hot research topic among the scholars. The "knowledge" search "homework burden" related literature themes amounted to 2982 whereas, the "outside training" related literature had 5459 sightings. The main aim of the researchers was to create better education environment to realize quality education goals in policy implementation. The relevant information revealed that the study of education burden theme mainly includes: primary and secondary school teachers, quality education, burden policy, and other related aspects. However, the current "double reduction" policy has failed to achieve the ideal intended effect therefore, it is necessary to probe this aspect in-depth for better and effective policy implementation.

In July 1955, the MOE published instructions on reducing the burden of primary and secondary schools pointing out that it is necessary to improve the relevant teaching material, optimize teachers' teaching level and increase extracurricular open activities to improve students' mental and physical abilities.

Subsequently, the MOE continuously issued following policies to reduce the burden of education: on May 4, 1964, the Report on Overcoming the Overburden of Primary and Secondary Schools and Improving the Quality of Teaching, the Issuance of the Provisions on Reducing the Overburden of Primary School Students on May 11, 1988, and the Emergency Notice on Reducing Student Overburden in Primary Schools was issued on January 3, 2000. From 1985 till date more than 50 times several policies were issued on Education burden-reduction indicating the concern of Chinese government but many of the policies ended as anticlimactic.

The government is forward-looking and sincere in its efforts to reshape the education ecology and promote education equity from various perspectives, including institutional setting and policy protection. After the issuance of the "Double Reduction" Opinions, the MOE has held eight press conferences and briefings related to "Double Reduction" to explain the policy objectives, its significance and initiatives, and held pilot city promotion meetings to take stock of progress in the work. In 2021, the MOE conducted a comprehensive investigation and cleanup of out-of-school training advertisements in which the market supervision departments around the country closed 2,119 cases of out-of-school training advertisements [2]. The data during the week from March 4 to 11, 2022, revealed that "double reduction" was one of the top hot issues of education in China's "two sessions" [3]. At one time it was debatable that whether the implementation of "double reduction" education policy is more crucial than the program itself, since the implementation of the policy involves many influencing factors, the most central of which is the front-line teachers in primary and secondary schools, who execute the education policy and function as drivers of educational change.

Therefore, is necessary to establish a conducive environment which will help teachers to internalize and externalize the “double reduction” policy and eliminate teachers’ cognitive and behavioral disorders against the policy to make it more effective. This is the only way to remove the teachers’ perceptions and behavioral barriers against the “double reduction” policy and make relevant adjustments in educational governance and the change the concept of talent training implied in the “double reduction” policy.

The measures taken abroad to reduce the burden of education have domestic reference significance. Primary and secondary schools in Finland emphasize stimulating students’ originality in the classroom teaching process. Meanwhile, the schools adopt the incentive education evaluation mode to timely optimize and adjust according to the curriculum needs of the students to nurture high-quality school leaders and teachers. The specific content of Japan’s “loose education” measures can be summarized as “two reduction, two reduction and one increase”, that is, to reduce the teaching content, reduce the teaching difficulty, shorten the teaching time, shorten the weekly class hour to achieve a complete Friday system, and increase comprehensive learning time. At the same time, the Japanese government uses advanced IT means to carry out online courses which is no longer limited to on-site teaching. Online teaching is more transparent than that in China and is cheaper than on-site teaching and its combination with the workload of primary and secondary school teachers and reduces the burden on teachers. The British solutions include: carrying out periodic workload survey, paying attention to the pertinence of the burden reduction policy, testing the implementation process and effect, and providing training to teachers. On the other hand, Germany and Japan governments strictly prohibit in-service teachers to conduct part-time classes. They classify in-service teachers into the civil service system and follow strict requirements for in-service teachers in accordance with the relevant civil service system. Most scholars in the United States advocate the control of intervention in off-campus training institutions.

Learnings from some foreign countries public school systems can improve the status of domestic teachers, cultivate teachers ‘professional ethics, constantly optimize the quality of education in public schools, increase the government’s support for public schools, so that students can fulfill their needs and demands as much as possible in school. We should closely adapt to China’s national conditions, absorb the essence of foreign ways and methods of education burden reduction, and improve the deficiencies in the implementation of China’s education burden reduction policy.

Therefore, it is important to investigate and analyze the awareness, attitude and implementation of the “double reduction” policy among the front-line teachers in primary and secondary schools in order to analyze the difficulties in the implementation of the “double reduction” policy in a targeted manner. The knowledge, attitude, belief and practice model (KABP model), focuses on the whole process from cognition to belief building to behavioral practice and provides a scientific perspective for analyzing the “double reduction” policy. This empirical study is a survey and analysis of the current situation of the frontline teachers’ knowledge and beliefs about “double reduction” in Chenzhou City, Hunan Province, based on the KABP model, which can lead to effective implementation of the “double reduction” policy.

2. Study Design of the KABP Model

Since the implementation of the "Double Reduction" opinion, the primary and secondary schools across the country have taken corresponding actions including teachers from Chenzhou City. The city has 1951 schools of all levels and types, including 29 secondary vocational schools, 44 general high schools, 263 junior high schools, 388 primary schools, and 75,253 teaching staff in 2020. The high school student population (including technical secondary school) is 745,066, junior high school 1,864,176 and 1,205,918 students (including graduates, undergraduates and current students) from various schools. The city's resident population has an average of 9.58 years of education for those aged 15 and above [4].

The consolidation rate of nine-year compulsory education is 98.5% and the gross enrollment rate of high school is 93.1% [5]. The strong promotional measures of the national "double reduction" policy in Chenzhou City resulted in lot of fruitful reforms in the areas of arrival and departure time of primary and secondary students, extracurricular training, homework management, etc. The burden of homework and extracurricular training on students is reduced, and both teachers and schools have supported and approved the "double reduction" reform. A conducive atmosphere of "reform" has gradually formed, with a certain degree of representation.

2.1 Research indicators and survey tools

The "double reduction" problem is reexamined by using "KABP" model to structure and measure the "double reduction" effect and necessary countermeasures. The first is the horizontal composition of the "double reduction" effect, including the four dimensions of the policy namely, students' arrival and departure time, students' homework, the participation rate of extracurricular tutoring institutions, and the impact on teachers. The horizontal structure of the "double reduction" policy includes four dimensions: students' time away from school, students' homework, the participation rate of extracurricular tutoring institutions, and impact on teachers. The second is the longitudinal structure of identity, including three levels of knowing, believing, and practicing, which is the "double reduction" effect of knowing, believing, and practicing: knowing is cognition and knowledge, which is the knowledge level of mastery and understanding; believing is trusting and agreeing, which is the positive construction of belief level; practicing is action, which is the behavioral level of practice correction [6]. A questionnaire was designed on "double reduction" based on the "Opinions on Double Reduction" and the preliminary interviews with some primary and secondary school teachers, with a cross-section of horizontal and vertical dimensions. There were 19 questions, including 3 basic relevant information questions, 13 single-choice questions, 1 multiple-choice question, and 2 open-ended questions.

2.2 Sampling and survey methods

In 2022, a stratified random sampling method was used to select teachers from 12 primary and 12 secondary schools in Chenzhou City (6 in urban areas and 6 in rural areas, 6 in secondary schools and 6 in elementary school), and the questionnaires were administered to 242 respondents through the Chenzhou compulsory education teachers' group, of which 10 invalid questionnaires were excluded the effective rate of the valid questionnaires was 95.9%.

2.3 Statistical methods

Excel software was used for real-time statistical input of the data, and SPSS22.0 software was used for statistical description. One-way ANOVA or independent sample t-test was used for the agreement of the "double reduction" effect in different regions, and $p < 0.05$ was considered as statistically significant.

2.4 Study hypothesis

H1: There is no significant difference in the awareness of the "double reduction" policy among primary and secondary school teachers between city and township areas in Chenzhou.
H2: There is no significant difference in the beliefs of the "double reduction" policy among primary and secondary school teachers between city and township areas in Chenzhou.
H3: There is no significant difference in the practices of "double reduction" policy between primary and secondary school teachers between city and township areas in Chenzhou.

2.5 Reliability test

Reliability is the consistency of the same object when same measurement is carried out in the same method, in other words it reflects reliability of the measured data. The scale adopts three aspects: the awareness, beliefs and practices of the "double reduction" effect. The measurement dimension includes the time when students enter and leave the school under the "double reduction" policy, students' homework, the participation rate of extracurricular tutoring institutions, and the impact on teachers. The Cronbach coefficient was used for the reliability analysis of each measured dimension, and the overall (Cronbach's Alpha) coefficient is 0.832 indicating high reliability. The Cronbach coefficient of all dimensions is above 0.7 indicating that the questionnaire has good reliability.

3. Empirical Results of the KABP Model

3.1 Awareness of the "double reduction" policy among primary and secondary school teachers

Table 1 shows that all primary and secondary school teachers in Chenzhou are totally aware of the "double reduction" policy namely, 67.5% and 61.2% in urban and rural areas while the rest exhibit general awareness. The chi-square test for independent samples shows that the chi-square value is 0.897 and the progressive significance is 0.344＞0.05 hence, the null hypothesis is accepted representing that there does not exist a correlation between the region and the degree of knowledge.

Table 1 Awareness of the “double reduction” policy among primary and secondary school teachers

Region	Number of surveys	Knowledge level (Ratio)			*Pearson's chi-squared test*	*P*
		More familiar	General Understanding	No knowledge		
City	160	67.5	32.5	0		
Township	72	61.1	38.9	0	0.897^a	0.344
Total	232	128.6	71.4			

Note: The expected count of a. 0 cells (0.0%) is less than 5. The minimum expected count was 24.83.

3.2 Degree of belief among primary and secondary school teachers in the “double reduction” policy

Table 2 to Table 5 show that the standard deviations of urban recognition and township recognition are 0.716 and 0.776 respectively in all dimensions indicating that the recognition is relatively stable. The proportion of overall support and firm support for the “double reduction” policy is higher between urban and rural areas, the proportion of general implementation and above is higher, the proportion of importance attached to the “double reduction” policy is higher, and the degree of agreement that the “double reduction” policy can develop sustainably is higher. However, there are statistical differences between urban and rural areas in the overall attitudes of primary and secondary schools towards the “double reduction” policy, primary and secondary school teachers belief “double reduction” policy ; and schools implementing the national “double reduction” policy and the “double reduction” policy. There are no significant differences in the implementation of the national “double reduction” policy, the degree of importance attached to the implementation of the national “double reduction” policy, or the attitude of getting sustainable implementation in primary and secondary schools.

Table 2 Attitudes of primary and secondary school teachers towards the “double reduction” policy

Region	Number of surveys	General Attitude (Ratio)				*Pearson's chi-squared test*	*P*
		Strong support	Support	Not much support	Against		
City	160	54.3	41.2	3.7	0.8		
Township	72	45.8	40.2	8.5	5.5	8.330^a	0.053
Total	232	100.1	81.4	12.2	6.3		

Note:a. 3 cells (37.5%), the expected count is less than 5. The minimum expected count was 1.55.

Table 3 Teachers’ opinions on implementation of “double reduction” policy

Region	Number of surveys	Implementation (Ratio)				*Pearson's chi-squared test*	*P*
		Very well implemented	Better implementation	Implementation of general	Poorly implemented		
City	160	33.8	40.6	20.6	5		
Township	72	27.8	34.7	34.7	2.8	5.517^a	0.147
Total	232	61.6	75.3	55.3	8.3		

Note:a.1 cell (12.5%), with an expected count of less than 5. The minimum expected count was 3.10.

Table 4 Importance of school leaders in "double reduction" policy implementation

Region	Number of surveys	Level of Importance (Ratio)				Pearson's chi-squared test	P
		Attention	More important	Not much important	No attention		
City	160	51.2	36.3	11.3	11.3		
Township	72	50	37.5	6.9	5.6	4.524[a]	0.237
Total	232	101.2	73.8	18.2	16.9		

Note: a. 2 cells (25.0%), with an expected count of less than 5. The minimum expected count was 1.86.

Table 5 Teachers attitudes towards the sustainable implementation of "double reduction" policy

Region	Number of surveys	Sustainable Implementation (Ratio)			Pearson's chi-squared test	P
		Able to	Hardly	Not sure		
City	160	62.5	28.1	9.4		
Township	72	62.5	33.3	4.2	2.189[a]	0.335
Total	232	125	61.4	13.6		

Note: The expected count of a. 0 cells (0.0%) is less than 5. The minimum expected count was 5.59.

3.3 Degree of implementation of "double reduction" policy in primary and secondary schools

3.3.1 Students' arrival and departure time under "double reduction" policy

Generally, the degree of implementation of the students' arrival and departure timings in urban and rural areas is obvious. The results of the survey show that students have a higher percentage of time of arrival between 7:30 a.m. and 8:00 a.m. and a higher percentage of time of departure between 5:30 p.m. and 6:00 p.m. There is statistical difference in students' arrival time in the morning, with the urban students arriving at school before 6:30 a.m. is more than the rural students and after 8:00 a.m. urban students arrived at school is significantly less than rural students (p = 0.000).For statistical differences between students leaving school in the afternoon, urban students leaving school was significantly lower than rural students before 5:30 PM; urban students leaving school from 5:30 to 6:00 PM were significantly higher than rural students (p = 0.000).

Table 6 Students' arrival time in the morning

Region	Number of surveys	Time of Arrival (Ratio)					Pearson's chi-squared test	P
		Before 6:30	6:30 to 7:00 pm	7:00 to 7:30	7:30 to 8:00 pm	After 8pm		
City	160	0	11.9	37.5	34.3	16.3		
Township	72	2.8	15.3	5.6	50	26.4	28.979[a]	0.000**
Total	232	2.8	27.2	43.1	84.3	42.7		

Note: a. 2 cells (20.0%), with the expected count of less than 5. The minimum expected count is.62; **p<0.05

Table 7 Students' departure time in the afternoon

Region	Number of surveys	Time of Departure (Ratio)					*Pearson's chi-squared test*	*P*
		Before 5:30 pm	5:30pm to 6:00pm	6:00 p.m. to 6:30 p.m.	6:30pm to 7:00pm	After 7:00 pm		
City	160	24.4	56.9	5	3.8	10		
Township	72	56.9	27.8	2.8	0	12.5	27.619[a]	0.000**
Total	232	81.3	84.7	7.8	3.8	22.5		

Note: a. 3 cells (30.0%), with the expected count of less than 5. The minimum expected count was 1.86; ***p*<0.05.

Tables 6 and 7 show that there is significant difference in the arrival time of urban and rural areas besides , there is difference in the students' departure time e. g,, the percentage of students leaving schools after 5:30 and 6 p. m. in the townships is only 56.9% while, it is 27.8% in the cities.

3.3.2 Students' homework situation under "double reduction" policy

It is noticed in the study that the primary and secondary school teachers systematically take students' homework. Table 8 shows that the amount of homework for students is limited to 1 to 2 hours daily barring few schools in both city and township schools. Training on the homework design is conducted in both urban and rural areas to more than 70%, whereas around 25% schools neglect it. The frequency of training sessions on the assignment design in both urban and rural areas is once a semester. However, around 15% of the schools in both types do not conduct training sessions. It is evident from Tables 8-10 that there is not significant difference in the frequency of training sessions on the homework design in all schools (p = 0.015), and the proportion of urban students doing homework for less than 1 hour per day is lower than that of rural students'; while there is significant difference in the amount of homework held and students' homework per day for example, about 20% of urban and township primary and secondary schools did not conduct training on homework design.

Table 8 Daily homework done by students

Region	Number of surveys	Workload (Ratio)					*Pearson's chi-squared test*	*P*
		No more than 1 hour	No more than 2 hours	No more than 3 hours	No more than 4 hours	No more than 5 hours		
City	160	35	39.4	16.9	8.1	0.6		
Township	72	52.8	37.5	9.7	0	0	11.952[a]	0.008**
Total	232	87.8	76.9	26.6	8.1	0.6		

Note: a. 3 cells (30.0%), with an expected count of less than 5. The minimum expected count is.31;***p*<0.05.

Table 9 Status of Primary and secondary school teachers on assignment design

Region	Number of surveys	Convening (Ratio)		*Pearson's chi-squared test*	*P*
		None	There are		
City	160	25	75		
Township	72	27.8	72.2	0.200[a]	0.655
Total	232	52.8	147.2		

Note: The expected count of a. 0 cells (0.0%) is less than 5. The minimum expected count was 18.62.

Table 10 Frequency of training sessions on assignment design

Region	Number of surveys	Training Frequency (Ratio)				*Pearson's chi-squared test*	*P*
		0 times a semester	Once semester	Once per year	Twice per year		
City	160	16.9	64.4	4.4	14.4		
Township	72	15.3	77.8	0	6.9	6.800[a]	0.028**
Total	232	32.2	142.2	4.4	21.3		

Note: Note: a. 2 cells (25.0%), with an expected count of less than 5. The minimum expected count was 2.17; ***p*<0.05.

3.3.3 Participation rate of extracurricular training institutions in the "double reduction" policy

Table 12 shows that after the implementation of the "double reduction" policy, the participation rate of extra-curricular tutoring institutions in Chenzhou City has dropped significantly and more than 60% of teachers believe that the participation rate of students in extra-curricular tutoring is hardly 10%. This shows that the students' learning burden has reduced after the implementation of policy. Tables 11 and 12 show that there is a significant difference between urban and rural school students' participation in extracurricular tutoring before the implementation of the "double reduction" policy (p = 0.000), and urban students' participation is significantly higher than rural students indicating that the such training in urban areas is gradually decreasing.

Table 11 Students' participation in extracurricular training institutions before implementation of the "double reduction" policy

Region	Number of surveys	Participation (Ratio)					*Pearson's chi-squared test*	*P*
		Around 10%	Around 20%	Around 30%	Around 40%	About 50%		
City	160	11.9	15.6	27.5	15	30		
Township	72	40.3	22.2	12.5	4.2	20.8	32.019[a]	0.000**
Total	232	52.2	37.8	40	19.2	50.8		

Note: a. The expected count of0 cells (0.0%) is less than 5. The minimum expected count was 8.38; ***p*<0.05; the time cut-off points before and after the implementation of the "double reduction" policy in Chenzhou is defined on July 24, 2021.

Table 12 Students' participation in extracurricular training after implementation of the "double reduction" policy

Region	Number of surveys	Participation (Ratio)					*Pearson's chi-squared test*	*P*
		Around 10%	Around 20%	Around 30%	Around 40%	About 50%		
City	160	60	19.4	10	5	5.6		
Township	72	68.1	15.3	8.3	2.8	5.6	1.691[a]	0.782
Total	232	128.1	34.7	18.3	7.8	11.2		

Note: a. 2 cells (20.0%), with the expected count of less than 5. The minimum expected count was 3.10.

Table 13 reveals that only 40% teachers believed that students' participation in extracurricular training would be beneficial to their physical and mental development. However, teachers were relatively conservative or on wait-and-watch mode about the physical and mental development benefits from students' participation in extracurricular training. At the same time, it is noticed that there was no significant difference among the expectations of urban and rural teachers.

Table 13 Teachers' opinions on students' participation in extracurricular tutoring institutions under the "double reduction" policy

Region	No. of surveys	Impact (Ratio)				*Pearson's chi-squared test*	*P*
		Useful	Not beneficial	Harmful	Harmless		
City	160	42.5	40.6	3.8	13.1	0.752	0.855
Township	72	45.8	41.7	2.8	9.7		
Total	232	88.3	82.3	6.6	22.8		

Note: a. 1 cell (12.5%), an expected count of less than 5. The minimum expected count was 2.48.

Thus, in general, it can be concluded that primary and secondary schools are able to implement the basic requirements set by the national "Double Reduction" policy to a significant extent with positive changes in the arrival time of students in the morning and in the leaving time of students in the afternoon, students' participation in off-campus training, the daily homework for students, and the frequency of training sessions on the home assignments design. At the same time, majority (60%) of the school teachers are skeptical about sharp dropping in extracurricular training institutions attendance after implementation of the "double reduction", which subsequently would reduce students' learning burdens. Whereas, only 40% of teachers believed that students' participation in extracurricular training institutions would have a positive impact on students' physical and mental development. However, there still remain some problems which need to be addressed. First, there are inherent differences among urban and rural schools in the implementation of the "double reduction" due to demography. In urban schools, the implementation efforts are significantly higher than those in rural schools in terms of students' arrival and departure time, students' participation in off-campus training before and after the implementation of "double reduction", training on the homework design, and the frequency of training sessions on homework design. The daily homework is higher in urban schools than in rural schools, so it is recommended that the "double reduction" rules should take into account the differences between urban and rural areas. Second, the "quantitative change" is greater than the "qualitative change". The "quantitative changes" in the time of arrival of students in the morning, the training on the homework design, and the frequency of training sessions on homework design is more obvious in all schools. However, the magnitude of the "qualitative changes" in students' afternoon departure time, students' participation in off-campus training before and after the implementation of "double reduction" is relatively small, which indicate that the promotion of "double reduction" should pay attention in improving quality, effectiveness and persistence.

4. Conclusions and Recommendations

The results of the present survey of 242 primary and secondary school teachers in Chenzhou City, Hunan Province, indicates that the "double reduction" policy has been implemented for more than one year, and the overall recognition, importance and support from school teachers is high. At the same time, China's regional differences, imbalanced development of urban and rural areas and schools demand top-level design and coordination of education policies to promote the orderly operation of the education system. There are significant differences between urban and rural areas in terms of awareness, recognition, and implementation of the national "double reduction" program as narrated below.

4.1 Knowledge level differences between urban and rural school teachers

It is noticed that the overall awareness of the "double reduction" policy among primary and secondary school teachers in Chenzhou is significant. However, urban primary and secondary school teachers are more aware of the national "double reduction" policy than their rural counterparts. Therefore, education authorities and schools should continue to increase teacher training on the "double reduction" policy with an appropriate focus on rural teachers.

4.2 Attitudes and beliefs differences between urban and rural areas teachers

It is observed that majority of the primary and secondary school teachers are positive and optimistic about the meaning and effects of the "double reduction" policy on students' healthy growth and support the policy. However, specific attention needs to be given on: many primary and secondary school teachers do not have positive expectations about the impact of "double reduction" on student' learning. Urban school teachers believe in the national "double reduction" policy more than rural school teachers, and nearly half of the rural teachers have a negative attitude towards the "double reduction" policy in general. Therefore, education authorities and school management should focus on the "double reduction" work in townships by strengthening, guiding, supporting to recognize the "double reduction" work in townships， so as to enhance the teachers' belief in the "double reduction" policy.

4.3 Quantitative and quantitative changes in student burden reduction

The "quantitative change" in the implementation of the "double reduction" policy is obvious and is attributed to better implementation of basic requirements of "double reduction" in terms of student's arrival time, increase in training sessions and homework checks and sharp dropping in the proportion of students participating in training in extracurricular tutoring institutions after the implementation of "double reduction", which helped to reduce students' learning burden. However, the magnitude of the "qualitative change" in the time students leave school in the afternoon and their participation in out-of-school training is relatively small. Therefore, to promote the "double reduction" improve the quality" has to be the starting point for action otherwise, it may lose the fundamental and "the original will cannot be inverted" [7]. The report clearly points out that "high-quality development is the primary task in comprehensively building the modern socialist country. " Compulsory education as a basic strategy is the process of national modernization, "quality change" is mandatory in the development of compulsory education [8]. Therefore, the present study suggests that Chenzhou primary and secondary school teachers should focus on "grasping lesson preparation". Under the background of "double reduction", teachers should study students' learning situation and teaching materials and their current level through "grasping lesson preparation", pay attention to the recent arena of students' development, and give full play to the leading role of teachers as well as students' role. Secondly, through "grasp the classroom" concept, teachers have to pay attention to streamline the classroom content on the basis of grasping the classroom. The classroom is the core place for students to learn, and "grasp the classroom" is an effective way to highlight the position of students, stimulate their enthusiasm for active participation, guide students to carry out inquiry learning, and improve efficiency of classroom teaching. Subsequently, "grasp the homework" will guarantee improvement in teaching efficiency. Therefore, primary and secondary school teachers in Chenzhou city should consider the homework content situational. Homework integrates and assimilates classroom knowledge into specific applied situation. It is closely linked to student' learning life, and forms a typical case study or project by processing in the real world. Secondly, the homework design operation needs to consider the difficulty progression. The assignment is no longer a repetition of knowledge without difficulty gradient, but an advanced structured design and the design assignments need to be considered in a variety of ways. It is no longer a unified written homework, but an organic combination of homework with students' housework, communication, interview, social experience and so on, to form mathematics in life, so as to realize the diversity of homework methods [8]. At the same time, "double reduction" is not just a slogan to improve quality, the implementation is quite challenging and systemic project that requires synergy between government, school and family in participation and common construction. If there is no multi-party collaboration, it will be difficult for teachers to receive effective guidance and support, and they are bound to become more anxious and helpless under the demand of reducing the burden and improving the quality. Therefore, can we find a way to promote the collaborative practice of schools, government, teachers, and parents, so that the "double reduction" policy can be transformed from a confident "knowledge" to a reliable "belief" and finally to an effective "action"? This is the key point to reduce the burden and improve the quality. On the one hand, from school point of view, we should devote ourselves to integrate high-quality education and teaching resources, strengthen teaching and research guidance on curriculum and teaching; empower high-quality development with information technology, and build efficient classrooms [9], On the other hand, the education department should work with women's federations and other departments to run parenting schools or online family education guidance platforms and promote the construction of community family education guidance centers and service sites [10], Only by integrating the power of collaborative governance inside and outside the school building, developing a new ecology of lifelong education for primary and secondary school teachers, improving the ability of teachers to know, believe and act on the "double reduction" policy, and adjusting expectations for academic progress and development based on each student's personality, we can provide emotional and technical support to teachers to confidently and boldly reduce students' academic burden and fully implement the "double reduction" policy.

The study has following limitations: on the one hand, the sample range and sample size are insufficient, the fillers are subjective and random, and there are defects in objectivity and conscientiousness; on the other hand, the variable design of the personal information of the respondents is not comprehensive, and the factors such as gender, age, discipline, educational background, teaching experience, age and marriage of the respondents are not fully included and analyzed. The limitation of this study and its discussion can also serve as an important consideration for further scientific research on “double minus” problem in the future.

Acknowledgements

Authors are grateful for the grants from the Educational Science Planning Project of Hunan Province [XJK20BGD034] to support this study.

References

[1] General Office of the CPC Central Committee General Office of the State Council. Opinions on further reducing the burden of homework and off-campus training on students at the compulsory education level [EB/OL]. (2021-07-24) [2023-04-16]. http://www.moe.gov.cn/jyb_xxgk/moe_1777/moe_1778/202107/t20210724_546576.html.

[2] Jin-peng Huai. Report of the State Council on Effectively Reducing the Burden of Excessive Homework and Out-of-School Training and Promoting the Comprehensive and Healthy Development of Students at the Compulsory Education Stage [EB/OL] (2022-10-28) [2023-04-16]. http://www.npc.gov.cn/npc/c30834/202210/71d4c80e30824de1b157a26f3f1be50b.shtml.

[3] Li Kang, Hao Huang. Contributing to the promotion of educational equity and quality improvement [N]. China Teachers’ Journal, 2022-03-16(001).

[4] Chenzhou City Local History Compilation Office. Chenzhou Yearbook [M]. Beijing: Unity Press, 2021: 259-358

[5] Chenzhou City Bureau of Statistics. Chenzhou City Statistical Bulletin of National Economic and Social Development in 2021 [EB/OL]. (2022-04-06). [2023-04-16]. http://www.czs.gov.cn/html/zwgk/57666/57702/content_3417759.html.

[6] Ping Zhang. An empirical study of mainstream ideological identity education of college students--a theoretical perspective based on the model of knowing, believing and acting [J]. Educational Academic Monthly, 2021, 344(03): 92-97+111.

[7] Bing Zhang, Tianjun Chen. Expedient implementation: the practical logic of after-school services in schools [J]. Educational Development Research, 2021, 41(Z2): 50-58.

[8]Ninghai Chen. Promoting school-based training for teacher development [J]. Fujian Basic Education Research, 2022, 159(03): 17-19.

[9] Juan Ni, Zhen Chen. The value of bridging and clarifying the relationship between “reducing burden” and “improving quality”-after the implementation of the “double reduction” policy the risk of “not reducing but increasing” after the implementation of “double reduction” policy [J]. Journal of Education, 2023, 19(01): 126-137.

[10] Hai-ping Xue. The practical dilemma and optimization path of “double reduction” policy implementation [J]. Educational Development Research, 2022, 42(22): 3.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230196

Advantages and Strategies of Ideological Education in University Physics Course

Ailing ZOU [1], Huanlin LV, Lulu ZHAO, Lin WANG and Ying SHI
Department of Basic Education, Dalian Polytechnic University, Dalian, Liaoning, China

Abstract. Curriculum Ideology and politics plays an important role in promoting the development of higher education and training talents. This paper reported the unique advantages of the ideological education in university physics curriculum. Based on the practice of teaching physics, some strategies for the development of the concealed ideological education are proposed. Taking physics learning as the entry point to carry out ideological and political education, we can cultivate student's moral cultivation and patriotism. Therefore, it is not only beneficial to improve the interesting and ideological nature of the class, but also to enrich student's political spirit and build a complete and noble personality.

Keywords. University physics course, Ideological Education

1. Introduction

In 2016 National University Ideological and Political Work Conference, it was clearly put forward that "we should make good use of classroom teaching as the main channel, and the ideological theory course should be strengthened, enhance the affinity of ideological education, meet the needs and expectations of students' growth and development. Meanwhile, other courses should keep a good responsibility field, so that all kinds of courses and ideological theory courses can go hand in hand, forming a synergistic effect"[1]. In the process of human's pursuit of truth and exploration of the unknown world, physics shows a series of scientific world views and methodologies, which profoundly affect human's basic understanding of the material world, way of thinking and social activities [1]. The university physics course based on the basic knowledge of physics is a significant compulsory basic curriculum for students of science and technology majors. It is necessary to conduct ideological education in university physics curriculum [2]. For example, physics experiment is the link between theory and practice. University physics course can improve students' interest, mobilize their enthusiasm for independent learning, independent thinking and exploration and research [3]. At the same time, it can cultivate students' practice and invention. The physics class and physics experiments can integrate the curriculum with the ideological education well [4].

This paper explores the use of physics teaching as a carrier to mine the ideological and political materials of physics knowledge. We could build a curriculum ideological

[1] Corresponding Author, Ailing Zou, No1. Qinggongyuan, Ganjingzi Dist., Dalian, China; E-mail: zouailing_@126.com.

and political resource platform by improving the ideological and political awareness and ability of professional teachers.

2. The advantages of ideological education in university physics course

2.1. Physics course can cultivate students' scientific spirit

Physics is a science that studies the movement of matter and its laws. It can imperceptibly promote students to develop a correct outlook on the world, values and life [5]. Physical laws help students to establish basic knowledge of the scientific world, master objective laws, and form a scientific way of thinking. Through physics learning, students can understand fully the nature of the world. Meanwhile, physics course guide students to think deeply and research, and cultivate students' scientific spirit of seeking truth from facts and innovation.

2.2. Display of frontier knowledge of physics, enhance students' patriotism, make implicit ideological education logical

Course is the main channel for activating and guiding ideology. We can introduce some cutting-edge physical knowledge, such as electromagnetic ejection and magnetic levitation, into the physics classroom. It not only breaks the mystery of high-tech products, shorten the distance between students and the frontiers of science and technology, but also stimulate students' enthusiasm for scientific and engineering invention and patriotism[6].

3. The Strategy of Developing Ideological Education in university physics class

3.1. Clarifying the physics course objectives of ideological education in university physics

Universities should cultivate people who possess scientific literacy, social responsibility, national honor [7]. It is indicated that higher education should not only provide professional knowledge for student, but also give them guidance and attention in the spiritual level. Ideological education should be integrated into the whole university physics curriculum, especially in classroom demonstration experiment. We could cultivate students' thinking methods and the scientific belief of science. Meanwhile, China's scientific and technological development achievements are integrated into classroom demonstration experiments, which inspire students' emotion of patriotism.

3.2. Establishing the Ideological Resource Bank of University Physics courses

We should establish the ideological resource bank of university physics course [8]. The ideological elements mainly include: socialist core values, education of China's excellent traditional culture, craftsmanship spirit, national defense education as well as "FAST", "5G technology", "high-speed rail", "electromagnetic ejection", and "large aircraft", which are the world's leading characteristics in the field of engineering technology in

China. Take the course demonstration experiment as the practice carrier, develop relevant classroom discussion, situational teaching and other ways to implement ideological education.

3.3. Integrate ideological education into teaching process

Introduction of new lessons is the first procedure of teaching. Since the class has just started, the students' attention can't focus on the lesson. If we integrate ideology resources into the new lessons time of introducing the new lessons according to the knowledge content of this class, it can not only play a role in gathering people's minds, creating a good teaching atmosphere, but also increase students' interest and enthusiasm in learning. For example, when explaining Bernoulli equation, we introduced the demonstration experiment of blowing table tennis floating with a hair dryer. This experiment result is contrary to the expectations of the students. At this time, we paid special attention to the opportunity to carry out the ideological and political education of "practice is the only standard for testing truth".

According to the actual situation of knowledge content, the ideological elements should be properly integrated, so that knowledge elements can become an organic whole. The history of physics is also the history of the development of human cognition. In the process of knowledge teaching, the background and process of knowledge generation should be timely introduced. It is not only beneficial for students to better understand and master knowledge, but also can increase their knowledge, experience the hardships of the scientific research process, and cultivate the scientific spirit of fearing difficulties and daring to explore. When teaching physical knowledge, we should pay attention to contact philosophy. Physics is the basis of natural science. Marxist "historical materialism" and methodology are based on the cognition of natural laws. Therefore, ideological education in university physics courses has unique advantages. When teaching the microscopic particle wave particle duality of quantum mechanics, students should be guided to think dialectically from the perspective of philosophy. In teaching relativity, students should be guided to discuss the law of cause and effect in philosophy and follow the trend to carry out the education of Marx's historical materialism.

3.4. Carrying out Ideological Education in Classroom by Demonstrating Experiments and hands-on experiments

Firstly, the basic knowledge of physics and Marxist dialectical historical materialism are organically integrated in university physics class by classroom demonstration experiments and hands-on experiments. The experiment points out the opposition and unity of contradictions, which enhance students' dialectical view of problems [9]. Secondly, through hands-on experiments, students are encouraged to apply physical knowledge to solve practical problems, and focus on frontier physics knowledge, which could cultivate students' observation and thinking ability, practical ability, and scientific spirit of teamwork. Otherwise, the experiments demonstrate the frontier technology research achievements in China, which enhance students' self-confidence and national pride, and strengthen students' belief in contributing their own strength to the great rejuvenation of the Chinese nation. Finally, Ideological Education could focus on inspiring students to inherit and carry forward their glorious traditions with the great artisans, learn their excellent qualities of hard work and striving for the rejuvenation of China, and stimulate students' spirit of innovation and exploration. At the end of the section on resolution of optical instruments, we introduce the 500m aperture spherical radio telescope "China Tianyan" located in Pingtang County, Guizhou Province, so that students can understand the application of the resolution formula.

3.5. Promoting interactive communication between teachers and students

Teachers should target students' physical and mental development stages according to the teaching rules of course knowledge, and help students choose the most suitable communicate style. In the teaching process, it is only the most basic requirement for students to acquire physics knowledge. The real purpose of teaching is to enable students to apply the learned physics principles to solve practical problems, apply theories to practice, and ultimately promote students' abilities. Therefore, in the teaching process, we must not only start from the textbook, but also extend the classroom. In the teaching process, we also focused on the relationship between students and teachers. Effective communication and exchange are the key factors to completing the goal of ideological education [10].

In sum, "Ideological education" is a teaching concept, university physics curriculum is the main carrier of ideological education. The ideological education of curriculum is not an isolated and stiff in the teaching process, teachers should naturally, smoothly and effectively spread the corresponding content and concept of "ideological education of curriculum" to students [11]. We can achieve a win-win situation of rational knowledge learning and perceptual ideological education. Otherwise, teachers themselves should constantly improve their political quality, personal accomplishment and patriotic feelings, rather than inconsistent empty words and platitudes. This kind of curriculum can really move and affect students. The proportion of time for ideological education in the whole classroom teaching should be appropriate. The ideological and political time of a class should not exceed 15 minutes, and it is better to intersperse it in the teaching process.

4. Conclusion

Ideological education is a new requirement put forward by the CPC Central Committee for all kinds of courses in universities in the new era. Under this situation, the aims for all universities are cultivating students' moral. It's important to achieve ideological education throughout the education process. In the university physic class, we should clarify the ideological objectives of the physics curriculum, establish appropriate ideological elements based on university physics, and achieve moral cultivation in the imperceptible influence. Recently, the teaching reform of ideological education in university physics is still at the exploratory stage. In the future teaching practice, it is necessary to constantly enrich the teaching elements of ideological education, and enhance the education quality.

Acknowledgements

This work was supported by the Science Research Special Fund of Liaoning Province (No. J2020057).

References

[1] "Big Data Analytics for Cyber-Physical System in Smart City", Springer Science and Business Media LLC, 2020.
[2] Yu Rong, et.al, "Implementing ideological and political education in university classroom teaching", 2010 International Conference on Artificial Intelligence and Education (ICAIE), 2010.
[3] Y. Liu and H. Zhang, "Research on online teaching in colleges and universities," in Chongqing Higher Education Research, vol. 8, 2020.
[4] Y. Zhao, "Theory and practice of online teaching of History in high school," in Continue Education Research, no. 5, 2020.
[5] Chunyan Wei. "Reform of College English Teaching from the Perspective of Ideological and Political Education", Lifelong Education, 2020
[6] F. Sun, D. Ji, and X. Liu, "Research on mixed online and off-line teaching modes based on Wisdom Tree platform---A case study of Automated production line installation and debugging'", in Wireless Internet Technology, no.1, 2020.
[7] Y. Xia and Y. Yang, "Reflections on online course 'Analytic Chemistry' in the context of an epidemic", in University Chemistry, vol. 35, 2020.
[8] R. Xue, H. Liu, J. Zhang and X. Wu, "Reflections on online teaching modes of the course of Plant Physiology," in Education Teaching Forum, no. 35, 2020.
[9] X. Zhao, T. Gao, P. Ge and J. Zhang. "Reconstruction and design of mixed teaching within the context of Internet," in China Electric Power Education, 2019.
[10] C. Xie, S. Li and F. Wang. "An analysis of the challenges faced by college teachers and students in the context of an epidemic,' in Statistics and Management, no. 8, 2020.
[11] X. Zhan. "A brief analysis of the development of online education in colleges and universities in post-epidemic period," in Science and Technology Vision, no. 23, 2020.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230197

A Corpus and Computer-Assisted Translation-Based Study on English Translation of Intangible Cultural Heritage Terms

Menglei Lv [a,1] and Liping Jiang [b]
[a] *Software Engineering Institute of Guangzhou, Guangzhou city, Guangdong province, China*
[b] *Guangdong Mechanical & Electrical Polytechnic, Guangzhou city, Guangdong, China*

Abstract. The translation of intangible cultural heritage (ICH) texts presents significant challenges due to their rich cultural connotations and unique regional characteristics. To address this, the author of this study utilized Corpus and Computer-assisted Translation (CCAT) to assist in the English translation of ICH during the creation of a parallel corpus. Three patterns of CCAT-based English translation of ICH were proposed: machine translation, human-machine interactive translation (combining human and machine translation), and human translation. The implementation of these patterns can improve the efficiency and quality of ICH translation, promote terminology standardization, and facilitate the exportation and transmission of ICH.

Keywords: Corpus and Computer-assisted Translation (CCAT); Intangible cultural heritage; Chinese-English Translation; Intangible cultural heritage terms

1. Introduction

China's rich and extensive culture has a history of thousands of years, and with the implementation of the "the Belt and Road" initiative, promoting cultural "going global" has become an excellent opportunity for cultural development and dissemination [1]. Intangible cultural heritage (ICH) is a crucial component of China's traditional culture and its translation serves as a cultural window for foreign exchange. However, due to the distinct regional and dynamic characteristics of ICH, its English translation presents significant challenges [2].

The author uncovered a wide variety of problems in the existing English translation corpus of ICH texts during the creation of a parallel corpus. For example, in the introduction of local cultural ICH, there are phenomena of excessive transliteration. For instance, "英歌" in Jieyang, Guangdong Province was translated as "Yingge" and "雷剧" in Zhanjiang, Guangdong Province was translated as "Leiju". This simple

[1] Corresponding Author, 1531447813@qq.com, Guangdong Mechanical & Electrical Polytechnic, Guangzhou city, Guangdong, China.

transliteration fails to take into account the acceptability of the target language readers and could not accurately convey the cultural connotation of ICH [3]. Among them, the most severe and influential problem on the international dissemination of ICH is the lack of term standardization, namely, the problem of confusion in the translation of names. The lack of uniformity in translated names in ICH texts indicates the translators' failure in adopting modern translation technology to standardize their translation and also implies low translation efficiency [4].

Machine translation continuously updates the processing paths and means for natural language. The introduction of artificial intelligence has induced machine translation into the era of neural network translation, greatly improving translation speed and efficiency [5]; Applying the computer-assisted translation system to E-C translation and C-E translation has advanced the overall quality to a passing level or above, and enables users to basically achieve the communicative purpose [6]; The construction of a bilingual parallel corpus has provided the best platform for language research and translation research [7]. By constructing and applying a corpus for ICH translation research, researchers can more deeply explore and recognize the laws and characteristics of ICH texts at the linguistic level, and hence summarize objective and universal principles and methods for ICH translation [2]. Using modern translation techniques, including corpus, translators can not only achieve terminology unification, but also greatly improve the accuracy and efficiency of translation [4]. Corpus and Computer-Assisted Translation (CCAT) [1] provides an excellent way for translators to improve the efficiency of intangible cultural heritage translation.

Based on the theoretical concept of CCAT (Corpus and Computer-Assisted Translation), this research takes the English translation of introduction text of the national ICH as an application case to explore the English translation pattern of ICH under the CCAT platform. The corpus used is a national Chinese-English (C-E) bilingual corpus of ICH, which contains three million aligned words of ICH corpus, covering national ICH corpus texts from more than ten cities across the country [2]. Translators can search the corpus for authentic translation examples to assist in translation practice, and can also use translation assistance platforms, such as Snowman CAT software and Trados software, to assist their translation, aiming to achieve the goal of unifying terminology and improving translation efficiency.

2. Literature review

Corpus and Computer-Assisted Translation (CCAT), in which Corpus mainly refers to bilingual corpus, mainly specialized in bilingual corpus and includes online corpus as well, while the computer-assisted part specifically refers to CAT software [1]. The future CCAT system will be built as an online platform and connected to idioms, henceforth every user of the system, such as the translator, the proofreader, the reviewer, and the translation researcher, will belong to the sub platform of the system, sharing the public corpus, terminology, and memory database [4].

Laviosa [8] pointed out that corpus research has been thoroughly integrated into translation research since the early 1990s, and with the development of corpus research, it has also exerted a significant impact on the cultivation of our translation thinking abilities and translation teaching. The symposium on "Corpus Use and Learning to Translate" held in Italy in 1997 opened a new chapter in the application of parallel corpus to interpreter training/translation teaching [9]; Since then, foreign scholars such as Zanettin (1998:616-630) [10], Bowker (1998:631-651, 2003:169-183) [11][12], Barlow

(2000:106-115) [13], Bernardini (2004a: 15-34, 2004b: 97-112) [14][15], and Pearson (2007:15-24) [16] have begun to study the application of corpus in translation teaching and make teaching attempts. Corpus based translation research started relatively late in China, but it has developed rapidly. In 2003, Ye (2003:41-44) [17] explored the issue of C-E translation at the lexical level by building an English Chinese parallel corpus. Wang (2004:27-32) [18] is recognized as a pioneer in exploring the application of corpus in translation teaching, initiating a boom in domestic attempts to use corpus in translation teaching.

The application of corpus and computer-assisted translation systems into translation research has caused heated debate in the translation community. The author conducted a search on CNKI and the result is listed as follows: Based on "Corpus" and "Translation", 9556 results has been shown; Based on "Computer-Assisted" and "Translation", 2282 results has been displayed; Based on "Corpus and Computer-Assisted Translation (CCAT)", 254 results has been exposed. For example, Shen and Zhang (2018:254) [19] took the snowman CAT software as an example to explore the application of computer-assisted translation software based on parallel corpus in translation teaching. They constructed a comparative experimental model and elaborated its application in translation teaching, and then proposed an optimized innovation model based on the results of the experiment; Zhu and Chen (2015:52) [20] developed a computer-assisted translation teaching platform and applied it in translation practice teaching. According to visualized research data, few studies have combined computer-aided translation with the ICH translation pattern for research. However, translation research and teaching applications based on corpus and computer-assisted translation software have become a research trend, so it has certain exploration significance applying corpus and computer-assisted translation in ICH translation research.

3. Three patterns of the CCAT-based English translation of ICH

CAT Translation Memory technology stores and classifies diverse and massive language data in the form of a technical corpus and corpus and then implement the automatic matching during translation. Meanwhile, the main popular CAT tools, such as Trados and DejaVu, also provide the function of convenient text content capture [21]. CAT software mainly has the following three functions: 1) Managing the translation project, facilitating terminology standardization. 2) Avoiding repetitive translation thus to improve the translation efficiency. 3) Establishing a matching mechanism hence to improve the translation efficiency.

Based on the self-built national ICH corpus, utilizing existing original texts and translations, the research team has formed a translation memory and then updated it in real time. When the translator is translating ICH, the computer-assisted translation system will automatically search and match the same or similar translation resources (such as phrases and sentences) in the translation memory, and then provide reference for translators. Therefore, translators only need to focus on the translation of new content, which can not only avoid unnecessary repetitive work, but also can ensure the unity of the translation of the same content.

Owing to the fact that national ICH translation materials contain amounts of expressions with rich cultural connotations, dealing with culture-loaded words simply with CAT model has revealed numerous drawbacks. Therefore, this study will propose three patterns of the CCAT-based English translation of ICH: machine translation, human-machine interactive translation (combining human and machine translation:

human-assisted translation and machine-assisted translation), and human translation, based on the terms and special expressions with rich cultural elements in ICH texts.

3.1. Machine translation

Machine translation has a relatively high translation speed and efficiency, and has been widely used in translation fields with high degree of informative texts. Translation software can be adopted directly to translate simple and objective declarative sentences with little cultural connotation and plain language in some ICH texts.

Example 1: 五羊传说(from national level ICH of Guangdong)

CAT: The legend of the five rams

The English translation shown in Example 1 is simple and clear, and can be easily understood by target language readers, so it can be adopted directly. Then, by searching for the English translation of the term "五羊传说" in the network corpus, the author has further verified that the term is professional.

3.2. Human-machine interactive translation

Machine translation and human translation have their own advantages and disadvantages respectively. The advantage of machine translation is its fast speed and high efficiency, but some machine translations may be rigid, obscure, and fail to accurately convey the cultural connotations of ICH. Although human translation is more accurate than machine translation to some extent, it cannot efficiently handle a large amount of translation tasks or those emergent tasks. The human-machine interactive translation pattern combines both translation methods, complementing each other's advantages. By making full use of parallel corpora of ICH, network corpora, and computer-assisted translation systems, translators can conduct human translation to maximize translation effectiveness.

The human-machine interactive translation includes two main operation modes: human-assisted translation and machine-assisted translation.

3.2.1 Human-assisted translation

With today's technological advancements, machine translation still cannot completely replace human translation. To achieve a balance between translation quality and efficiency, human-assisted translation is necessary. Human-assisted translation should be adopted to translate ICH texts with complex sentence structures and little cultural connotation but simple and objective declarative sentences. Human-assisted translation mainly involves text processing before translation and post-editing after translation. Text processing before translation includes adding a subject to sentences without one based on context, removing redundant vocabulary, and replacing special vocabulary. Post-editing includes selecting and replacing words in the translated text and arranging sentence structure, etc.

Example 2: 历史上由唐代宫廷狮子舞里脱胎而来,五代十国之后,随着中原移民的南迁,舞狮文化传入岭南地区.(National level ICH)

Text processing before translation: A subject has to be added to the first sentence according to the context for there is no subject in it; The chunk "历史上" has been removed for the similar connotations of it with the other chunk "脱胎而来", which was then replaced by another chunk "演变而来" for the later one is more easily been read by the machine. Thus the original text has been edited and translated as:

CAT: It evolved from the palace lion dance in the Tang Dynasty. After five dynasties and ten countries, with the migration of immigrants from the Central Plains to the south, the lion dance culture was introduced into Lingnan.

Post-editing: As is recommended in the network corpora, it is more acceptable to translate "宫廷狮子舞" as "court lion dance". According to Chinese history, it is more appropriate to translate "五代十国" as "the Five Dynasties and Ten States (907 B. C.~960 B. C.)". Firstly, "五代十国" is a proper noun, hence it should be capitalized. Secondly, according to Chinese history, it is not agreeable to translate "十国" as "ten countries". Furthermore, by searching into the network corpora, the author has found that in 80% corpus, "五代十国" is translated as "Five Dynasties and Ten States". Finally, adding the time annotation after the expression "五代十国" can help the target language readers to understand the history and culture during such a period. Moreover, the sentence structure has been adjusted due to the different sentence structures of Chinese and English.

Post-edited version: It evolved from the court lion dance of the Tang Dynasty. After Five Dynasties and Ten States (907 B. C.~960 B. C.), the culture of lion dance was introduced into Lingnan as the people in the Central Plain immigrated to the south.

3.2.2 Machine-assisted translation

ICH texts contain massive terms with cultural connotations, thus, as previously mentioned, English translation of ICH terms is deemed as a difficulty. For example, there are phenomena of excessive transliteration. For instance, "佛山十番" was translated as "Foshan Ten Scenes" and "英歌" of Jieyang was translated as "Jieyang Yingge". Such simple transliteration fails to consider the acceptance of target language readers, and cannot accurately convey the cultural connotations of ICH. While machine translation of these terms is also not ideal. According to the literature, "佛山十番" is a type of folk instrumental music, which belongs to gong and drum music. Therefore, "Foshan Percussion Music" is a more appropriate translation. As is recorded in the literature, there are several types of "英歌", including "英歌" (Puning, Jieyang), "潮阳英歌" (Shantou), and "甲子英歌" (Shanwei), all of which have different characteristics for they originated from different regions. Therefore, the adjusted translations of "Puning Ying-Ge Dance", "Chaoyang Ying-Ge Dance", and "Jiazi Ying-Ge Dance" are more appropriate and easier for target language readers to accept.

To ensure the terminology standardization, avoid repetitive translation of the same content, save resources, and ensure the accuracy of translations, translators can adopt machine-assisted translation with the help of computer-assisted translation (CAT) tools, ICH parallel corpora, online translation, and translation dictionaries. Translators can verify their translations of terms via the network materials and Chinese-English parallel corpora, and then added the verified terms to the ICH Chinese-English corpus for real-time updating, hence to standardize the translation of ICH terms through the translation memory retrieval and matching.

3.3 Human translation

Human translation can be flexibly adapted to specific contexts and grammatical requirements. For special expressions with distinct cultural connotations, human creativity in translation can make it easier for the target readers to understand. Based on the source content, human translators can create new expressions in the target language by combining transliteration and free translation, which can not only help the readers understand the source content better but also have a deeper impact on the readers. In fact, it is more conducive to the dissemination of Chinese culture.

Example 4: 主要套路有"采青", "高台饮水", "狮子吐球", "踩梅花桩"等. (National level ICH)

CAT: The main routines include "picking green", "drinking water from high platform", "lion spits ball", "stepping on plum blossom pile", etc.

The original text contains numerous expressions with rich cultural connotations. Machine translation mostly adopts the method of free translation to deal with these expressions, and the part of speech of theses expressions in the machine translation are not uniform. By searching the relevant information on the internet and then studying the translations in the ICH corpus, the author has confirmed the translations of the terms and then stored them into the translation memory to update the ICH translation corpus. Additionally, the CAT translation reads still too rigid, thus it is adjusted as follows:

Human translation: They will also represent the set patterns of "Caiqing" (picking the green), "Gaotaiyinshui" (drinking on a hathpace), "Shizituqiu" (spitting small balls) and "Caimeihuazhuang" (stepping on the staggered pilings).

In terms of the obscure Classical Chinese rich in the complex traditional Chinese cultural elements, human translation is an effective translation strategy.

4. Conclusion

In conclusion, this paper highlights the benefits of using modern translation technologies, such as corpora and computer-assisted translation systems, on the CCAT platform to integrate human and machine translation for the English translation of ICH. The proposed translation patterns, including machine translation, human-machine interactive translation (combining human and machine translation), and human translation, provide effective solutions for different types of ICH texts. The study demonstrates that the CCAT-based approach can significantly improve translation efficiency and accuracy, ensuring the unification of terminology in translations.

Moving forward, this research can be applied to other translation fields that require creative translation of culture-loaded words, ambiguous language, and professional terms. The next step of the research is to apply the ICH corpus in translation teaching and construct an ICH English translation teaching platform system under the CCAT platform. The proposed new translation teaching modes will be tested through empirical research on translation major students, offering valuable insights and guidance for research on translation teaching in the new era.

Acknowledgement

This research was supported by the Humanities and Social Sciences Research and Planning Fund Project of the Ministry of Education of China (2021) (Grant No. 21YJA740016); the Key Scientific Research Project (Special Scientific Research Project) from Guangdong Mechanical and Electrical Polytechnic (Grant No. YJZD2021-47).

References

[1] Yu, J. Translation research on world heritage under the CCAT platform. Journal of Qiqihar University (Philosophy, Social Science Edition), 2018,(02), 135-137.
[2] Jiang, L. P. Creation and application of the Chinese-English parallel corpus of intangible cultural heritage. China's Intangible Cultural Heritage, 2022,(02), 118-124.
[3] Jiang L, Al-Shaibani GKS, Yang F, Cheng M and Huang M (2022) The metonymic mechanism of English translation of Chinese intangible cultural heritage terms from the perspective of cognitive psychology. Front. Psychol. 13:957485. doi: 10.3389/fpsyg.2022.957485
[4] Jiang, L. P. Exploring translation strategies of intangible cultural heritage from the perspective of cross-cultural communication: Taking the English translation of Guangdong intangible cultural heritage as an example. Journal of Qingyuan Vocational and Technical College, 2022,15(01), 69-74.
[5] Jiang, L. (2022). Factors influencing EFL teachers' implementation of SPOC-based blended learning in higher vocational colleges in China: A study based on grounded theory, Interactive Learning Environments, DOI: 10.1080/10494820.2022.2100428
[6] Yang, W. D., & Fan, Z. R. Analysis of post-editing in machine translation of scientific and technological texts. Shanghai Translation, 2021,(06), 54-59.
[7] Li, F. Q. A comparative study on the quality of English-Chinese machine translation based on neural networks. Shanghai Translation, 2021,(04), 46-52.
[8] Wang, K. F. . The design and development of Chinese-English parallel corpus. China Foreign Language, 2012,9(06), 23-27.
[9] Laviosa, S. Corpus-based translation studies: theory, findings, applications. Amsterdam: Editions Rodopi B.V.,2002.
[10] Xiong, B. Research on translation teaching mode based on Chinese-English parallel corpus. Foreign Language World, 2015,(04), 2-10.
[11] Zanettin, F. Bilingual comparable corpora and the training of translators. Meta, 1998, 43(4), 616-630.
[12] Bowker, L. Using specialized monolingual native-language corpora as a translation resource: A pilot study. Meta, 1998, 43(4), 631-651.
[13] Jiang, L., Liang, X. (2023). Influencing factors of Chinese EFL students' continuance learning intention in SPOC-based blended learning environment. Educ Inf Technol. https://doi.org/10.1007/s10639-023-11734-4.
[14] Barlow, M. Parallel texts in language teaching. In Botley, S.P., McEnery, A.M., & Wilson, A. (Eds.), Multi-lingual Corpora in Teaching and Research (pp. 106-115). Amsterdam: Rodopi Bv Editions,2000.
[15] Bernardini, S. Corpora in the classroom: An overview and some reflections on future developments. In Sinclair, J.M. (Ed.), How to Use Corpora in Language Teaching (pp. 15-34). Philadelphia: John Benjamins,2004.
[16] Bernardini, S. Corpus-aided language pedagogy for translator education. In Malmjær, K. (Ed.), Translation in Undergraduate Degree Programmes (pp. 97-112). Amsterdam: John Benjamins,2004b.
[17] Pearson, J. Using parallel texts in the translator training environment. In Zanettin, F., Bernardini, S., & Stewart, D. (Eds.), Corpora in Translator Education (pp. 15-24). Beijing: Foreign Language Teaching and Research Press,2007.
[18] Ye, C. Q. The application of self-built corpus in translation teaching: Classroom design for the translation of the Chinese and English versions of "A Dream of Red Mansions". Foreign Languages and Their Teaching,2003, (03), 41-44. DOI: 10.19716/j.1672-4720.2003.03.009.
[19] Wang, K. F. The use of bilingual parallel corpus in translation teaching. Foreign Language Electronic Teaching, 2004,(06), 27-32.
[20] Shen, Y., & Zhang, L. The application of computer-aided translation software based on parallel corpus in translation teaching: Taking the XueRen CAT software as an example. In Research on the Development and Innovation of Foreign Language Education and Translation 2018,(Vol. 7, pp. 254-257).
[21] Zhu, Y. B., & Chen, J. L. (2015). Theoretical conception and engineering practice of computer-aided translation teaching platform based on corpus. Foreign Language Electronic Teaching, (04), 52-57.
[22] Zhao, H. (2019). English translation of cross-border e-commerce product copywriting with human-machine interaction. Chinese Science & Technology Translators Journal, 32(01), 26-28+8.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230198

International Academic Influence of Sport Science Based on Scientometrics and Visualization

Nan SU[1]
International Studies School, Sichuan University, Chengdu 610064, China

Abstract. Taking 194 papers cited by Web of Science in China Sport Science and 176 citing articles as research object, use scientometrics methods and visualization analysis to make the time zone, core author, main institutions, research topic of cited literature. Clarify the citations of papers in Sport Science by Web of Science, to explore international academic influence of Sport Science. Overall, the international academic influence of Sport Science has been continuously promoted, which has become an important channel for the international community to understand the achievements of sports science in China.

Keywords: Sport Science; international academic influence; Scientometrics; Visualization

1.Introduction

Modern big data analysis tools are rapidly evolving, and scientometrics theory and tools are gaining traction in various fields of society, fueled by the emerging technological revolution of artificial intelligence and information interaction. Scientometric methods based on the logic of intelligent big data algorithms have become more important and active in promoting innovative discoveries in modern social science research. Academic journals play an important role in spreading academic ideas and strengthening information communication by serving as an important platform for displaying academic research results [1, 2]. This research focuses on the combined application of scientometric methods and modern social science research, and exploring the interactivity, coupling, mapping, and complex relationships of modern social science research via big data analysis.

The Chinese sports science journal "Sports Science" is included in the international high-level sports literature database, and it has a significant impact on sports science research both at home and abroad. Sports Science papers cover a wide range of topics, including sports sociology, sports economics, sports medicine, school physical education, and sports history, demonstrating hot topics and cutting-edge trends in Chinese sports science research. Scholars have discussed the evolution of Sports Science from various perspectives such as citation, journal level, and fund support. However, there is still a gap in research focusing on the academic impact of Sports Science from an international

[1] Corresponding Author, Nan SU, Sichuan University, China; Email: sunan2008_2010@163.com

perspective, which is not conducive to increasing the international impact of Chinese sports science journals and limits the application of modern information technology in the field of social science research. This paper attempts to compensate for Sports Science's lack of an international research perspective through a scientometric analysis of the citations of papers published in Sports Science by the Web of Science database and its visual presentation, which objectively reflects the journal's panoramic picture, academic value, and international influence.

2. Data sources and research methods

Scientific knowledge mapping is a relatively new research method in scientometrics and informetrics. It is rapidly being used in modern social science research because it can display the information structure of related fields in the form of a map and reveal knowledge sources and evolutionary laws. The mapping of scientific knowledge is primarily based on a special clustering operation of subject words, from which it derives the nodes and maps of high-frequency nodes, clustered knowledge groups, and high intermediary centers of a specific topic, as well as the hot spots and research frontiers, and its main mapping tools include Citespace, VOSviewer, SATI, and Ucinet.We finally got 194 articles cited by WOS database after literature search and data cleaning after selecting the Web Of Science database search platform, selecting the option of "Cited references search," and entering the English title of China Sport Science or China Sport SCI in the field of cited works.

Following the initial de-drying of the data using Notepad+ software, the bibliometric software HistCite, Sati, and the visualization software VOSviewer were imported to perform a multidimensional metrological analysis and visualization of the cited and sourced literature of Sports Science in the WOS database in order to comprehensively explore the international citations of Sports Science and its international academic influence[3]. Co-word analysis and visualization analysis are the primary bibliometric methods employed in this paper. The purpose of the visualization analysis is to demonstrate the complex relationship of research topics in the form of graphs and images through the quantitative calculation of literature elements, focusing on the key issues and mutations of the research object, and providing an important support for understanding the research field from its essence. It is a valuable resource for understanding the research field[4].

3. Analysis of cited literature

3.1 Time distribution

The citation frequency is an important index for evaluating the research level of literature, and it is an objective reflection of academic journals' knowledge dissemination ability and academic influence [5]. Statistics revealed that the WOS database cited 194 papers in Sports Science, totaling 337 citations and an average of 1.74 citations per paper. Since 2003, the number of Sports Science papers cited by the WOS database has increased dramatically, with two peaks in 2016 and 2020, indicating that the international attention and influence of Sports Science has grown but is not stable.

3.2 Highly productive authors and core institutions

Table 1 lists the top ten most prolific authors and core institutions in the Sports Science cited literature. Representative scholars of the cited literature in Sports Science include Feng Lian-shi, Hong Ping, Tian Ye, and Zhang Zhongqiu from the Institute of Sports Science of the State General Administration of Sports, Wu Yandan from Fujian Normal University, Lin Xianpeng from Peking University's School of Management, Zhang Lin from Shanghai Institute of Physical Education's School of Economics and Management, and Yao Gangyan from Wuhan Institute of Physical Education. The Institute of Science of the State General Administration of Sports, Fujian Normal University's School of Sports Science, Beijing University of Physical Education, Tsinghua University's Department of Physical Education and Research, Shanghai Institute of Physical Education, and others are among the key institutions cited in the literature. These institutions not only play an important role in Chinese sports science research, but they also actively promote international sports research cooperation and academic exchanges.

Table 1 The top 10 prolific authors of cited literature in Sports Science

No.	Author	Number	Unit
1	Feng Lian-shi	7	Institute of Sports Science, State General Administration of Sports
2	Hong Ping	5	Institute of Sports Science, State General Administration of Sports
3	Tian Ye	4	Institute of Sports Science of the State General Administration of Sports
4	Wu Yandan	4	College of Sports Science, Fujian Normal University
5	Lin Xianpeng	3	School of Management, Beijing University of Sports
6	Zhang Zhongqiu	3	Institute of Sports Science, State General Administration of Sports
7	Zhang Lin	3	School of Economics and Management, Shanghai Institute of Sports
8	Yao Gangyan	3	Wuhan Institute of Sports
9	Si Qi	3	Department of Sports Science, College of Education, Zhejiang University
10	Xiaoping Chen	3	Department of Physical Education and Research, Tsinghua University

3.3 Highly cited literature

Table 2 shows the list of highly cited literature in Sports Science cited by the WOS database, among which there are eight papers cited ≥ 4 times. The highly cited literature spans the years 2000 to 2012, but the average frequency of citations is low. The average citation frequency is far lower than that of comparable foreign journals, indicating that Sports Science's international academic influence needs to be increased [6, 7]. The majority of the authors of the highly cited literature are well-known sports science scholars. The highly cited literature covers a wide range of topics, including macro-level content such as the history of sports development, competitive sports, and the development of the sports industry, as well as micro-level reflections on national physical fitness monitoring, exercise energy expenditure, and meditation training.

Table 2 Highly cited literature in Sports Science (cited ≥2)

No.	Author	Title	frequency	Issue
1	Tian Yupu	A Retrospective and Reflections on the Shifting Strategic Focus of Sports Development in New China in the Past 60 Years	7	2010/01
2	Han Xiuying, Jiang Chongmin	Mathematical Modeling of the Logical Test Method of National Physical Fitness Monitoring Data in 2005	6	2009/03
3	Li Yinan, Wan Ping	Analysis Report on the Results of Implementing the Student Physical Fitness Standard at Peking University	5	2005/02
4	Si Qi	Path analysis among psychological factors affecting the stage change of exercise behavior among college students	4	2006/08
5	Xiong Xiaozheng, Zheng Guohua	The Formation, Evolution and Reconstruction of China's Competitive Sports Development Model	4	2007/10
6	Dai Jiansong, Li Jing et al.	Extrapolation of energy expenditure for walking and daily physical activity	4	2006/11
7	Kang, Jin-Young	Effects of meditation training on post-training anxiety, state of mind, and heart rate recovery in American high school swimmers	4	2000/06
8	Liu, Weina	Revision of the Chinese version of the Simplified State Fluency Scale and Simplified Trait Fluency Scale	4	2012/12

4. Analysis of the cited literature

4.1 Temporal distribution

The cited literature refers to cutting-edge issues in a specific discipline or research field [8]. A total of 176 citations were obtained through data cleaning and careful checking, with 297 citations, averaging 1.69 citations per article. Figure 1 depicts the main citations published between 2007 and 2016. In 2007, papers in the WOS database began to cite Sports Science literature, and the number of citations reached 23 in 2013, indicating that Sports Science is becoming an important platform for the international sports science community to understand relevant research results in China.

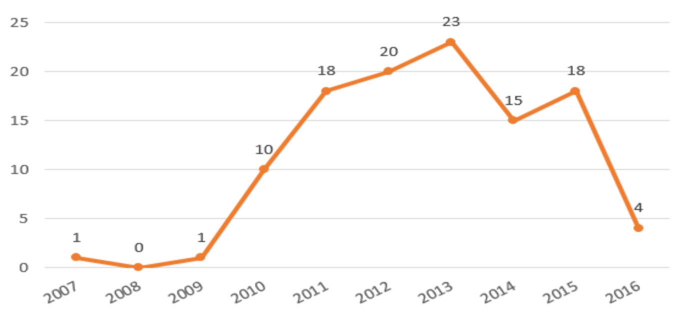


Figure 1 Time distribution of cited literature in Sports Science

4.2 Core authors, countries (regions) and core institutions

The core authors of the cited literature are analyzed to explore the main author groups that enhance the internationalization of Sports Science[9]. The study of the country/region and core institutions of the cited literature provides an understanding of the main radiation targets of Sports Science[10]. The core authors of Sports Science include Chunqing Zhang and Yanping Duan from Hong Kong Baptist University, Hui Zhang and Ruiji Liu from Shanghai Institute of Physical Education, Junpeng Pang from Wuhan Institute of Physical Education, and Cardinal, Bradley J. from Oregon State University (Oregon State University). The top countries/regions were China with 96 citations (85.2%), Australia with 6 citations (5.7%), USA with 5 citations (4.6%), and UK with 3 citations (2.6%). In terms of the institutional distribution of the cited literature, there were 9 institutions with the number of articles ≥3. Among them, Hong Kong Baptist University, Hong Kong Institute of Education, Curtin University (Australia), and Oregon State University (USA) have cited Sports Science, indicating that Sports Science is being recognized by research institutions in the USA, Australia, and Hong Kong, and related cooperation The exchange has been increasing.

4.3 Source Publications

Lecture Notes in Management Science, Advances in Education Research, Applied Mechanics and Materials, International Journal of the History of Sports, and other journals are cited in Sports Science. Twenty-nine percent of them were published in Lecture Notes in Management Science. The cited literature's source publications are related to international conferences such as the International Symposium on Sports Science and Engineering, the International Conference on Education and Sports Education ESE (International Conference on Physical Education), and so on.

5. Hot spot clustering analysis of cited and cited literature

Keywords reflect the fundamental themes of literature research and serve as an important foundation for delving into the core research content. Keyword co-occurrence analysis is a metric based on text content. The cited literature was imported into SATI and VOSviewer, and the high frequency words of literature co-occurrence were extracted to create a visual knowledge map of the cited literature and the cited literature (Figure 2).

Table 3 and Figure 2 depicts the co-occurrence mapping of the cited literature in Sports Science, demonstrating a high degree of aggregation among different research topics and the formation of interlinked and cross-linked research clusters. Six thematic clusters of cited documents are obtained when content analysis and map analysis are combined: First, take anxiety-background-combination-control group-experience-feeling-intervention-physical activity research. The contents include students' intelligence characteristics, influencing factors of teenagers' growth and development, quality characteristics of excellent athletes, tennis service technology, and so on. Second theme is citizen-competitive sport-culture-distribution-feature-history-management-policy-role, includes sports policy, cultural development, government role, sports education, sports policy, public sports service. Third is physical education, with theme of application-building-college-cultivation-current-situation-organization-teaching, includes sports strategy, physical education training for college students, physical education reform, educational policy and strategy, and so on. The fourth is sports medicine and training research, with theme of athlete-body function-correlation analysis-country-examination-height-index-male-region-technique, which includes athletes' physical characteristics, intellectual development, sports techniques. Fifth, research into social sports, like child-community-competition-family-mechanism-school-sports culture, including leisure sports development, policies, mechanisms, school education, community development.

A thorough examination of the subject clustering of cited and cited documents reveals some differences in the research contents of cited and cited documents, but the overall research topics are highly consistent. The cited documents in Sports Science contribute significantly to the knowledge base and research support for WOS cited documents, while cited documents expand existing subject content and research results while also significantly improving the cross-border, cross-cutting, and cutting-edge of sports science research.

Table 3 High-frequency Keywords in Sports Science (partial)

No.	Key words	frequency	No.	Key words	frequency
1	group	17	11	athlete	9
2	Management	15	12	Role	9
3	stage	15	13	sports consumption	8
4	Exercise	13	14	control group	8
5	physical activity	13	15	sports culture	8
6	College	12	16	competition	8
7	country	12	17	intervention	7
8	Index	11	18	validity	7
9	Change	11	19	body function	7
10	Research method	11	20	mechanism	7

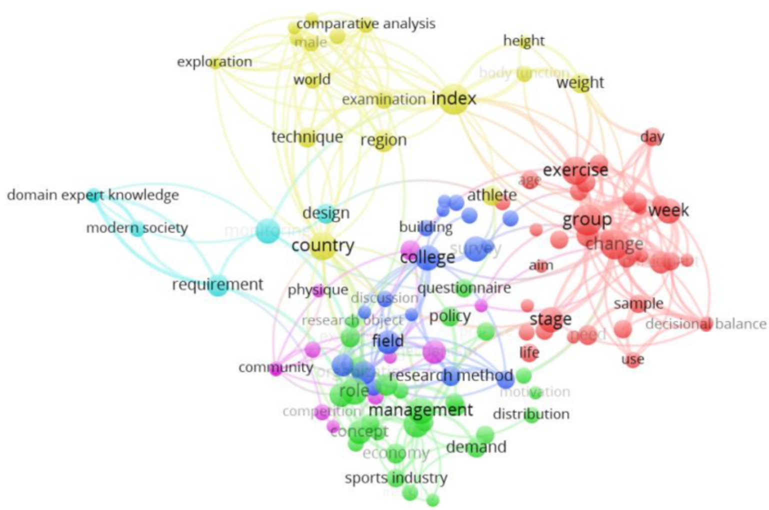


Figure 2 Keyword co-occurrence network of the cited literature in Sports Science

The measured keywords extracted from the literature show a relatively concentrated and hierarchical clustering, forming different research groups and reflecting the core content of research. The co-word network of the cited documents in Sports Science is analyzed, and it is discovered that the cited documents form several hot clusters. First, the research on the development path and mode of sports in China is centered on sports-sports industry-sports development-sports policy-sporting goods industry, with research topics including the development of sports in China, the direction of sports development, the formulation of sports policy, and so on. The second is basic theoretical research in physical education, with the core being athletes-psychological activities-exercise-sports physiology-physique-attention-national quality-burnout measurement, and the research topics including sports, sports training, athlete training, and so on. The third is competitive sports research, with the core being physical health-sports intervention-service technology-life care-disabled sports, and the research topics including sports, sports training, athlete training, and so on. Fourth, physical education research focuses on quality education, urban-rural differences in physical culture, school sports, university students, and other core topics. These include physical education for college students, physical education curriculum reform, sports research, youth training, and other content. Other hotspots include the Olympics, sports risk management, sports public finance, sports culture.

6. Conclusion

Artificial intelligence and interactive technology have propelled modern metrology into a completely new development space. Metrology technology and visualization methods play a significant role in promoting knowledge summarization, discovery, and interaction in social science research. Data mining of modern scientific research results is carried out in this paper using modern metrological tools such as citespace and vosviewer, which better present the application value of the research and obtain the basic conclusion that the papers published in Sports Science focus on the basic issues and main hotspots of Chinese sports science research, and respond to the realistic needs and the direction of national sports science development. Furthermore, an econometric analysis of Sports Science citations by the WOS database shows that the research level of the literature published in Sports Science has improved, the research content has become more diverse, and the gap between Sports Science and similar international authoritative journals is narrowing, and it has become an important channel for the international community to understand Chinese sports science research achievements.

In future research, it is necessary to further strengthen domestic and international collaborative research as well as inter author collaborative research, promote the cross integration of different research fields, enhance theoretical innovation in sports research, and promote the integration of Chinese sports discipline research with world leading sports research.

References

[1] Jiang CL, Liu S B, Ding K . Knowledge map of research hotspots and their evolution in China Journal of Science and Technology Research[J]. China Journal of Science and Technology Research, 2008(06):954-958.

[2] Bao G H. Analysis of the citation of papers in Chinese Science and Technology Journal Research by Web of Science database[J]. China Journal of Science and Technology Research, 2014(09):1186-1188.

[3] Zhang S-L, Cheng W-P. Theme visualization of national core journal conference papers from 2003-2012 based on co-word analysis[J]. China Journal of Science and Technology Research, 2014,25(7):936-940, 948.

[4] Eck N J V, Waltman L, Dekker R, et al. A comparison of two techniques for bibliometric mapping: Multidimensional scaling and VOS [J]. Journal of the American Society for Information Science & Technology, 2010,61(12):2405-2416.

[5] Hu Yifang. The relationship between citations and academic quality of academic papers[J]. Library Forum, 2015(05):56-59.
[6] Wang HJ, Sun JZ. Analysis of impact factor composition characteristics of international authoritative sports science journals[J]. Chinese Journal of Science and Technology Research, 2014(12):1519-1523.
[7] Li A Q. Comparative study on the evaluation of core sports journals at home and abroad[J]. Chinese Journal of Science and Technology Research, 2008,19(4):598-602.
[8] Kui JQ, Tang ZG. The disciplinary status of editing and publishing from the citation and citations of Science and Publishing in the past five years[J]. Science and Publishing, 2015(04):97-100.
[9] Guo Q, Zhao Ji . Research on academic influence of scientific and technical papers based on multi-attribute description of cited journals[J]. Library theory and practice, 2014(04):53-58.
[10] Li J, Guo XH, Jiang Q, et al. A preliminary study on knowledge mapping in safety science: the example of Safety Science journal data[J]. Chinese Journal of Safety Science, 2013(04):152-158.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230199

Path Exploration of Machine Translation Post-Editing Curriculum Construction in Vocational Undergraduate Education

Xiao XIAO[a], Hui SUN[a1], Jing ZHANG[b], Zhiliang LI[a]
[a] *Shanghai Technical Institute of Electronics & Information, China*
[b] *Shanghai Yizhe Information Technology Co., Ltd, Shanghai, China*

Abstract. This paper catches the key words about Vocational Undergraduate Education, machine translation and post- editing (MTPE) in the decade of 2013-2022, and analyzes their development trends, benchmarks through analyzing the current situation of the MTPE course for Applied English majors in Vocational Undergraduate Education and the necessity of the course setting with the help of VOSViewer software. Preliminary conception and exploration of the elements of the MTPE course design ideas, target positioning, course framework and course evaluation have been conducted, aiming at responding to the Ministry of Education's curriculum planning for the introduction to Applied English majors for Vocational Undergraduate Education--Machine Translation Post-Editing, which provides a referential path for talents training and curriculum transformation.

Keywords. Vocational Undergraduate Education; machine translation post-editing; curriculum design

1. Introduction

Vocational Undergraduate Education began with the needs of the times of industrial transformation and upgrading and labour market restructuring. Skilled technical talents cultivation featured with versatile, high-level, and application-oriented characteristics of technical skills personnel training requires the development of undergraduate and higher levels of vocational education. At the beginning of 2021, the introduction of the Administrative Measures on the Establishment of Undergraduate Level Vocational Education Majors (Trial) marked the official launch of China's Ministry of Education for the establishment of undergraduate majors in vocational colleges as well as the formal establishment of a national system for the major settings in Vocational Undergraduate Education. It also marks that Vocational Undergraduate Education has finally moved from research to comprehensive practice after years of verification. In September 2022, China's Ministry of Education published a vocational education profession profile (2022 revision). By comparing applied English majors offered by junior colleges and undergraduate ones, it is found that undergraduate education put high requirements for the traditional English majors transformation in terms of career orientation, objectives, professional competence requirements, professional courses, internship training and practice, professional certificates, and continuing courses. "Machine Translation Post-

[1] Corresponding Author, Hui SUN; E-mail: halleysunny@163.com.

Translation Editing (hereinafter referred to as MTPE)" was added to one of the core courses, and MTPE project training was put into the internship training session. Those changes highlighted the technical skill position, which bring about a clearer professional orientation for the field of linguistics in vocational education. The future career positions in language subjects are more clearly oriented, with translation as the main line of professional competence training and the increasingly urgent requirement for the development of technical translation skills.

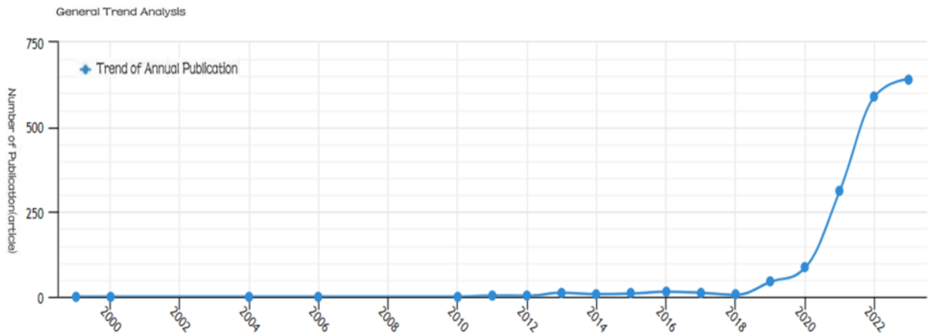


Figure 1 Developmental of Vocational Undergraduate Education in China

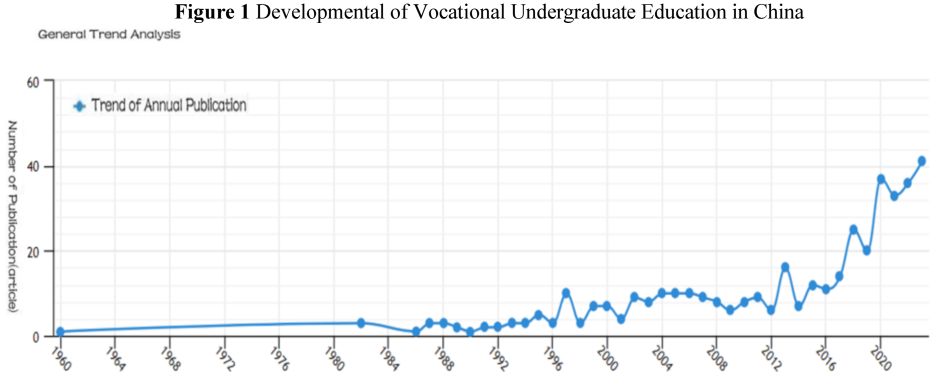


Figure 2 Developmental Trend of Machine Translation Post-Translation Editing

Inserting key words "Vocational Undergraduate Education", "Translation " in CNKI in Figure 1 and "Machine Translation Post-Editing" in CNKI in Figure 2, and Comparing both of them, whose vertical axis shows 'Number of Publication(article)', we can find the undoubted facts that:1. Vocational Undergraduate Education indeed has shown great booming in China, there are 1231 papers focus on its development, which has provided us with great opportunities and hope. However, when the key word MPTE is inserted, we find that though the general tendency is uprising, specific researches have shown weakness. There are only 372 papers that concentrate on MPTE research. There still exists great room for our Vocational Undergraduate Educational teachers to explore. Therefore, we resort to the VOSViewer to conduct a relatively thorough analysis of the two key phrases: MTPT and Vocational Undergraduate Education, the results have proved that more changes and researches are in great need.

In order to explore the research dynamics and research hotspots of the construction of the vocational undergraduate "Machine Translation Post-translation Editing" course in China and abroad, an exhaustive crawl of the core database on Web of Science (hereinafter referred to as WOS) during the decade of 2013-2022 on issues related to vocational undergraduate machine translation post-translation editing was conducted with the help of the visual analysis software VOSViewer, and an econometric analysis was done Through in-depth analysis, we can sort out the core disciplinary forces of international machine translation research; at the same time, based on author co-citation analysis, keyword co-occurrence and emergent word search, we can conduct a detailed study on the research hotspots, knowledge base and research frontiers in the field of international machine translation.

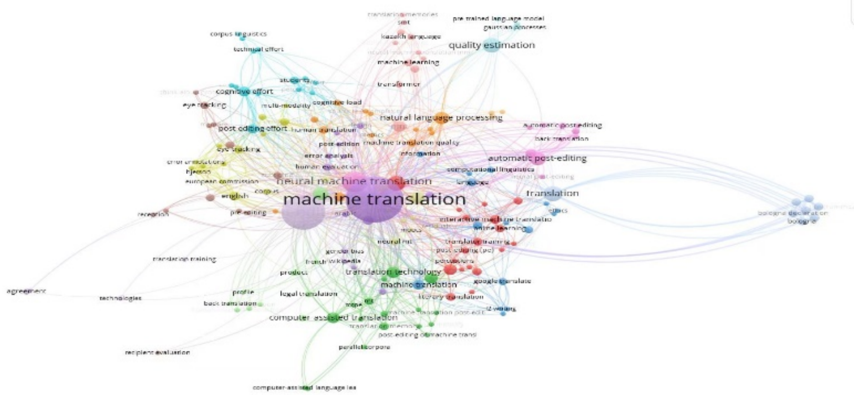


Figure 3 MTPE for Vocational Undergraduate Education (2013-2022)

The nodes and font sizes in Figure 3 are positively correlated with the volume of author publications; the larger the node, the more influential that author is in the field of international machine translation; the thickness of the lines between the nodes is also positively correlated with the volume of collaborative publications between institutions; the thicker the line, the closer the collaboration between scholars is reflected [1]. This section explores the issuing institutions of the selected corpus at a meso level, analysing in detail the research capabilities of different national research institutions in the subject area of post-translation editing for machine translation. In generating the plots, the threshold is set to 2.0 using VOSViewer and the network visualisation section presents the main research forces in current international machine translation post-editing research. By analysing Figure 3, the MTPE in this decade mainly involves: machine translation, translation technology, computer-assisted translation, interactive machine translation and so on. The author finds that the MTPE has shown a new trend of interdisciplinary, inter-university, inter-regional and transnational cooperation. The above keywords point to the general direction for the designation of post-translation proofreading courses for vocational undergraduates.

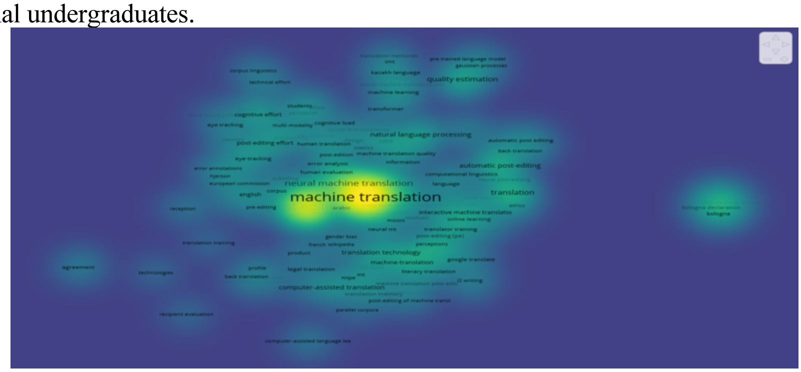


Figure 4 Density Visualization on MTPE between 2013-2022

When analyzing the visual density section of VOSViewer in Figure 4, we found that this decade's research on MTPE has focused on modules such as pre-translation proofreading, eye-tracking, brain-neural machine translation, natural language processing, corpus, and post-translation proofreading automation. From the nature of the current research institutions, most of those who have made significant progress in the field of MTPE are science and technology majors, and the results mainly present two major characteristics: firstly, it is easier to develop MTPE in multiple dimensions through the cooperation of institutions; secondly, the cooperation between schools and enterprises has led to the rapid transformation of research results. How to do a good job in teaching English-Chinese translation courses in relation to natural language processing and promote the deep development of machine translation post-editing in vocational undergraduate programmes is a key point to be explored by the front-line teaching staff of vocational undergraduate programmes.

2. Literature Review

Since its embryo in the 1930s, machine translation has been continuously evolved and updated. In 2015, neural network machine translation (NMT) was on the rise. Subsequently, well-known Internet companies such as Google, Microsoft, Sogou, and NetEase launched their own neural network machine translation engines, which quickly became popular language service products. Researches show that the output quality of neural network machine translation has improved significantly compared to that of statistical machine translation based on a bilingual parallel corpus translation model. However, there are still many common mistakes such as mistranslation, leaking translation, contextual inconsistencies, and inconsistent terminology, which are more prominent in more detailed vertical fields. Therefore, post-translation editing with human intervention has become an inevitable important part after the machine translation finishes the translation tasks.

According to the definition of the Translation Automation Users Association (TAUS), MTPE is a complex language process that requires editing the original text automatically generated by a machine translation system. Screen [2] states that MTPE usually uses as few human resources as possible to achieve higher efficiency than human translation. In National Standard (GB/T 40036-2021/ISO18587, 2017) [3], MTPE is defined on basis of machine translation results, with the aim of checking the accuracy and comprehensibility of machine translation, improving the text, enhancing the readability of the text, and correcting errors. The goal of post-translation editing is oriented to be recipients or consumers. Therefore, it'll be acceptable if the recipients or consumers or users think that the quality of the translation can meet their own requirements. Therefore, the judgment criteria of MTPE quality become more flexible. The types of MTPE can be generally distinguished by the degree of human intervention in machine translation. There are two main categories, including full post-editing and light post-editing. The outcomes of full post-editing should be accurate, understandable, and appropriate in style, using proper syntax, grammar, and punctuation, with the aim of producing translations that have the same effect as human translation products. Meanwhile, light post-editing may use the original machine translation results as much as possible to ensure that no information is added or omitted. The light post-editing task is to modify inappropriate contents and restructured incorrect or unclear sentences. Professional undergraduate talents focus on integrated technical skill management and service ability development for node positions in specialized fields. We propose that post-translation editing in the new MTPE course for undergraduate careers focuses on the proofreading of machine translation translations, which is closer to light post-translation editing in post-translation editing categories, that is, using the original product of machine translation as much as possible then to correct obvious errors and make the text easier for readers to understand. The focus of post-translation editing task is on revising the content of mistranslation, significant cultural differences, and reorganizing sentence structures. Light post-editing is post-editing with minimal human intervention, so the post-translation editing mentioned in this paper can be considered as light post-editing.

3. MTPE Course Design

Curriculum building is the foundation for the implementation of talent training and development. Liu Chengyou[4] proposed adhering to the guiding idea that Molding High Morals and Cultivating Talents Based on People Orientation, Attributes of Vocational Education and Higher Education should be integrated into the entire implementation process of Vocational Undergraduate Education, and Liu also emphasized that curriculum construction should adhere to the basic principles of “combining virtue and skill, placing equal emphasis on theory and practice”, “industry-education integration and work-integrated learning” and “student-centered”. Lu Jiancai[5] emphasized the need to follow the rules for cultivating skilled talents, construct a curriculum system and develop courses based on analysis of work tasks and professional abilities to refine the course development process and deepen the degree of school-enterprise cooperation. Yang Xinbin[6] proposed the establishment of a “technology-driven, product carrier, theory-practice integration, competency-based” curriculum system, implementation of a three-semester system with comprehensive practical courses, deep integration of industry and education, and further promotion of the “three teachings” reform, as an entry point and breakthrough for deepening content construction and improving the quality of teaching. Vocational Undergraduate Education should reflect the rank of undergraduate level on basis of adhering to professionalism. The course content requires deepening the knowledge study of technical theory, the cultivation of technological innovation ability, and the training of research and practical competence. From our point of view, the study conducted by Liu Chengyou[4] and Lu Jiancai[5] only provides the basic principles and guiding ideas for curriculum construction, no deep illustration and comments on course settings and strategies. Then, the course structure proposed by Yang Xinbin[6] highlights the professional characteristics of integrating industry and education, emphasizes the work procedures such as occupation, position, technology, skills, product and competence. Yang’s discussion is relatively comprehensive and systematic, which has the practical reference value to some extent, and can be applied to the design of the MTPE course. Therefore, we propose MTPE course ideas and framework featured with “Light post-editing technique driven, Applied translation product carrier, Translation theory-practice integration, Translation competence training targeted” (see Fig 5).

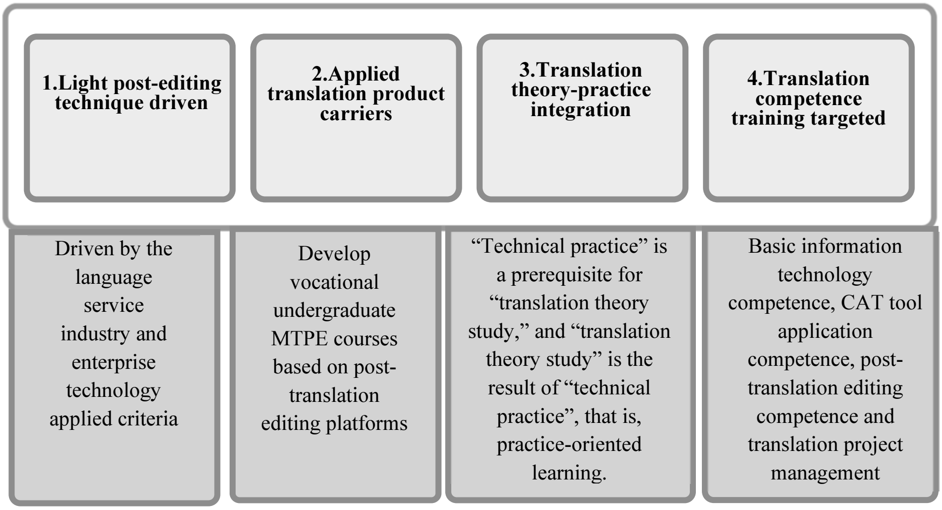


Figure 5 MTPE Course Design for Vocational Undergraduates

A. Light post-editing technique driven

MTPE in Vocational Undergraduate Education is light post-editing. Driven by the language service industry and enterprise technology applied criteria, MTPE course design should start from analyzing certain technical fields involved in the translators' work, analyze their relevant technical knowledge, then select and organize course content based on the work process and technological innovation process of the translators' post-translation editing practice. Meanwhile, the breadth and depth of technical knowledge should be laid emphasis on reaching the undergraduate level, that is, theoretical knowledge range may mainly cover from machine translation, post-translation editing to computer-aided translation, combining "vocational character" with level" while adhering to the professionalism.

B. Applied translation product carriers

When we develop vocational undergraduate MTPE courses based on post-translation editing platforms, some points need to be paid attention to. First, it is necessary to pay attention to select the typical post-translation editing platform, so that the platform can integrate the main technical knowledge of professional learning, and can be representative in the language service industry and enterprise production to ensure the integrity of translation technical knowledge learning. Second, it is essential to conduct practical process analysis based on post-translation editing platforms, which may include machine translation engine selection, post-translation editing strategy selection and terminological database application to ensure the matching between MTPE course and future work. Third, after post-translation editing practices, it is necessary to restructure course knowledge and competence system, design the relevant teaching and learning modules and form course framework and plan to ensure the inner logicality and systemization according to the whole MTPE task requirements.

C. Translation theory-practice integration

The integration of theory and practice in post-translation editing and proofreading of machine translations does not negate the importance of theoretical learning, but on the contrary, for Vocational Undergraduate Education, theoretical learning of translation should be of considerable breadth and depth; "technical practice" is a prerequisite for "theoretical learning of translation", which is the result of "technical practice", i.e. practice-oriented learning. The main position of practice in course of integration of theory and practice can better realise the unity of teaching objectives in terms of results and performance, and the unity of teaching implementation in terms of integration and contextualization.

D. Translation competence training targeted

The key element of MTPE course design is based on the core attributes of the subject. It is necessary to construct a translation technical competence model with Chinese characteristics that conforms to the foreign language talents training standards of Vocational Undergraduate Education. Wang Huashu and Wang Shaoshuang[7] believed that modern translators' translation competence included basic computer competence, information retrieval competence, CAT tool application competence, terminology competence and post-editing competence. Wang Shaoshuang and Qin Jianghua[8] built a translation technology knowledge framework composed of ten major sections, including basic knowledge, search technology, corpus technology, machine-assisted translation technology, machine translation technology, localization technology, terminology management technology, technical document writing, translation management technology, and other auxiliary technologies. Cui Qiliang[9] proposed that translators' technical competence should consist of information technology application competence, translation technology application competence and translation management technology competence. Based on the above analysis, we believe that the translation technical competence of vocational undergraduates should consist of four aspects which are basic information technology competence, CAT tool application competence, post-translation editing competence and translation project management competence (see Fig 6). Basic information technology competence includes basic computer competence and network search competence. CAT tool application competence includes the ability to use CAT tools, corpus technology, terminology management technology, quality assurance technology to complete actual translation tasks. Post-translation editing competence includes the use of machine translation technology, mainstream post-translation editing platforms and translation competence to complete post-translation editing tasks. Translation project management competence includes the ability to efficiently complete large-scale translation projects by using project management technology and team collaborative translation technology.

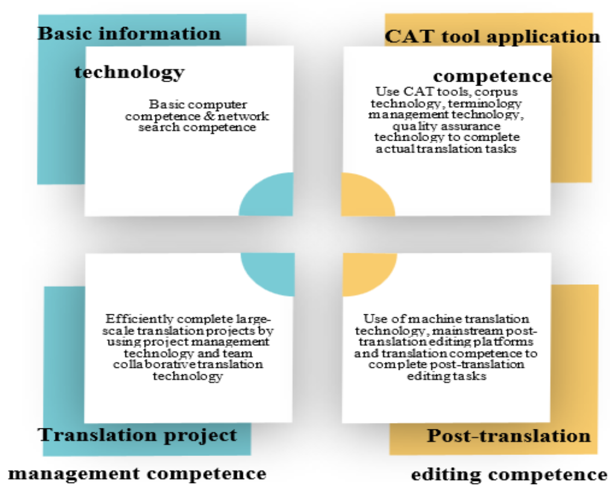


Figure 6 Translation Technical Competence of Vocational Undergraduate Education

4. MTPE Course Orientation

The Vocational Education Professional Profile (2022 Revision) [10] states that the vocational undergraduate applied English majors aims to train talents with high-level technical skills for comprehensively developing morality, intellect, physical work, beauty and labor abilities, mastering solid scientific cultural knowledge, English language foundations, basic translation theories and commonly used translation methods, knowledge of international trade business and relevant laws and regulations, owing strong language expression ability, on-site interpretation and data translation ability, foreign-related business processing, having the professionalism and information literacy of excellence, and being capable of excellence in business English translation, post-translation editing of machine translation, international business and foreign trade business. The national standard (GB/T 40036-2021/ISO18587, 2017) [10] proposed six competencies for post-translation editors, including translation competence, language competence and word processing competence by using source and target languages, competence to research, obtain and process information, cultural competence, technical and domain competence. Feng Quangong and Liu Ming [11] proposed a three-dimensional model of post-translation editing competence that includes cognitive dimension, knowledge dimension, and skill dimension, which is also in line with national standards for post-translation editing ability. Based on this, as a core professional course and internship training practice course for applied English majors, we believe that the general goal of MTPE course needs to combine the two dimensions which are English major standards for vocational undergraduates and national standards for post-translation editing competence in order to cultivate students to understand the development of the language service industry, market demand and post-translation editing competence, master translation project management workflows, and be proficient in using mainstream post-translation editing platforms to apply practical post-translation editing competence into machine translation projects (see Fig 7).

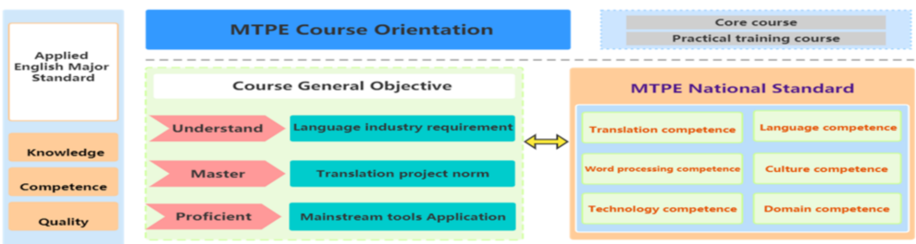


Figure 7 MTPE Course Orientation for Vocational Undergraduates

5. MTPE Course Framework

MTPE course can be conducted in various forms of teaching activities according to the characteristics of language major talents training model, current teacher status and development, teaching equipment and environment and depth of school-enterprise cooperation. Specifically, MTPE course can be divided into three model frameworks, that is, independent core course model, embedding translation course model, and internship training week course model.

A. Independent Core Course Model

Although post-translation editing seems to be editing machine-translated texts, in terms of the actual process of post-translation editing, the extension of post-translation editing should be widened to take pre-translation editing, post-translation editing, and general competencies into the overall systematic considerations. In detail, it should cover the cultivation of pre-translation text editing competence, post-translation text editing competence, and basic general competence. The independent setting of professional core courses generally requires 2 credits and 32 hours. For details, we can refer to Zhong Wenming [12] "Post-translation Editing Course Teaching Module" for adjustment setting. (see Table 1 for details)

Table 1 “Post-Translation Editing” Independent Core Course Model

Competence	Module	Content	Hour
Pre-translation text editing competence	Machine translation fundamentals	A brief history of machine translation; The concept of machine translation; The similarities and differences between MT and CAT;	2
	Common quality problems in machine translation	Common errors in current machine translation; Causes of machine translation errors; Methods for quickly locating common problems in machine translation	4
	Machine translation quality assessment	Indicators and methods for quality evaluation of machine translation	2
	Core skills for pre-translation editors	The purpose and role of pre-editing; The main strategies of pre-editing at the level of sentences, phrases and words (such as terminology)	4
	Mainstream post-editing tools/platforms	Online post-editing platform; Use of machine translation quality assessment tools; New interactive machine translation, integrated TM machine translation and automatic optimization of machine translation	2
Editing translated text competence	Principles of post-editing theory	The development history and industry status of post-translation editing; The general principles and business scenarios of post-translation editing	2
	Core skills of post-editing	Evaluation criteria of post-translation editing quality; Main strategies of in-depth post-translation editing; Main Strategies of Lightweight Post-translation Editing	4
	Post-translation editing project practice	Post-editing practices for documents in specialized fields (such as electronic information, technology, etc.)	6
Basic linguistic general competence	Project management	Project feasibility assessment, quotation and machine translation engine selection; Project time management/quality management/risk management	2
	Translation and intercultural communication	The role of translation ability in post-translation editing; Subject domain knowledge; Intercultural communicative competence	2
	General skills of translation technology	Basic text and image processing; Information retrieval; Office office basics	2

B. *Embedding translation course model*

In the early stages of course development, it is possible to explore post-translation content embedding models, so that traditional translation courses can be gradually transformed in the direction of language technology. Take the school-enterprise cooperative course between Shanghai Technical Institute of Electronics & Information (STIEI) and Shanghai Yizhe Information Technology Co., Ltd (YiCAT) as an example (see Table 2). The embedded content mainly includes six modules with 12 hours in whole. This content is taught by YiCAT staff who are sent to assist the STIEI in-class teaching activities. The first five modules focus on knowledge explanation to help students establish a knowledge system for machine translation and post-translation editing to cultivate students' post-translation editing awareness. The sixth module simulates "Post-translating Science and Technology News on the YiCAT Platform" as a case project simulation, uses the current YiCAT online translation management platform retriever to complete post-translation editing of texts in the field of news technology, then exports and submits task reports in the platform.

Table 2 Post-translation editing course embedding in "Occupational English Chinese Contrastive Translation Course"

Module	Content	Hour
A brief history of the machine translation development	Development, classification, working principle and application field of machine translation	2
Overview of machine translation quality evaluation	Common errors in machine translation; Classification, content, criteria and methods of machine translation quality evaluation	2
Call of machine translation in CAT tools	The use of mainstream CAT tools; The choice of different machine translation engines; The retrieval of machine translation in CAT tools	2
AI technology combined with machine translation	New interactive machine translation; Integrated TM machine translation; Automatic machine translation selection	2
Common error types and skills in post-editing	Post-editing competence; Common strategies and skills	2
Practical exercises for post-translation editing projects	Post-translation editing of technology news on the YiCAT platform	2

C. *Internship Training Week Course Model*

Internship training course is the feature of vocational education, which aims to help students improve their skill level, refine the quality of students' will, and serve students' all-round development. Internship training is generally conducted and run on a continuous teaching week. After preliminary exploration of the course implantation model, because of the epidemic influence in Shanghai in early 2022, STIEI Spring Teaching Courses were all transferred to online classes. The 2-week 60-hour MTPE course was adjusted to the "Friday+Weekends" intensive training model, completed in rotation over three weeks, with the aim of cultivating students' post-translation editing awareness and improving students' practical ability. (See Table 3 for details)

Table 3 MTPE Training Week Course

Module	Time	Content	Hour
Machine-aided translation platform cognition	Friday	Introduction to computer-assisted translation concepts; Main functions of computer-assisted translation tools; Translation memory	4
	Saturday	Introduction to online translation management platform; Introduction to YiCAT platform; YiCAT online practice (1): project creation, configuration, simulation translation and review	8
	Sunday	YiCAT online practice (2): Teachers distribute classroom practice assignments; students intend to be familiar with the process of project creation and editor use; students complete the training assignments independently	8
	Friday	YiCAT Training Homework Comments;	4

Simulated translation project cases		Project Management Knowledge Introduction; Translation Project Management Introduction	
	Saturday	Translation project management Task (1): team creation, team collaboration translation, translation submission and export	8
	Sunday	Translation project management assignments (2): Translation project process summary, project analysis report production, teacher-student communication and teacher comments	8
Post-translation editing project practice	Friday	Introduction to the development history of machine translation; Introduction to post-translation editing concepts and categories; Explanation of common machine translation errors and revision strategies	4
	Saturday	Post-translation editing assignments are distributed; Teachers demonstrate post-translation editing operations on the YiCAT platform; Students practice post-translation editing assignments	8
	Sunday	Teachers review the difficult points of post-translation editing based on student work reviews; Students revise the translation according to the review suggestions, and answer questions about the homework	8

6. MTPE Course Assessment

Whether it is independent core course model, embedding translation course model, or internship training week course model, MTPE course assessment should be focused on the general objectives and course orientation, emphasizing assessing students' technical knowledge reserves of translation technology, translation project management workflow, and practical competence to use mainstream post-translation editing platforms to conduct post-translation editing projects (see Table 4). A comprehensive evaluation can be made based on students' classroom attendance, group project work, pre-course preview, in-class discussions, post-course exercises, post-course preparation, in-class discussions, post-course practical exercises, and post-translation editing project completion quality.

Table 4 MTPE Course Assessment Form

Content	Translation technical knowledge reserve (Machine translation, post-editing, computer-aided translation, corpus, terminological database)	Translation project management workflow	Post-translation editing project practice
Proportion	15%	25%	60%

7. Conclusion

Under the guidance of Vocational Undergraduate Education, foreign language major settings are aimed at new business formats, new models, and new occupations, bringing great prospects for language service industry and MTPE market. The establishment of MTPE course for vocational undergraduate applied English major is still in its infancy. This paper proposes preliminary thought on the four basic elements of building vocational undergraduate MTPE courses. However, there is still a limit in the curriculum system, teaching equipment and resources, MTPE teacher team building, and integration of industry and education. It is our expectation that more experts and scholars may turn their attention and help applied English major and MTPE course development of Vocational Undergraduate Education, making contributions to the construction of China's modern vocational education system.

Acknowledgment

This study is supported by the project of Shanghai Vocational Education Association entitled Research on Digital Learning Mode in Vocational Education under the Epidemic Normalization Context (No.YB2022012, Project Master: Xiao Xiao)

References

[1] Chen C M. Predictive effects of structural variation on citation counts[J]. Journal of the American Society for Information on Science and Technology, 2012, 63(3): 431-449.

[2] Screen, B. “What effect does post-editing have on the translation product from an end-user's perspective?” The Journal of Missions Translation, issue 1, pp. 133-157,2019.

[3] GB/T 40036-2021/ISO18587:2017. Translation Service Post-Editing Requirement of Machine Translation [S].

[4] Liu Chengyou, Feng Liying, Zhao Feng. “Exploration on the Construction of Vocational Undergraduate Course System,” Journal of Hainan Normal University (Natural Science), issue 2, pp. 234-238, 2021.

[5] Lu Jiancai, Li Panpo, Jia Pengfei. “A Study on Undergraduate Level Curriculum Development in Vocational Education,” Education and Vocation, issue 23, pp. 102-106, 2021.

[6] Yang Xinbin. “Thinking and Exploration on Talent Training Mode of Vocational Undergraduate Education,” Research in Higher Education of Engineering, issue 1, pp. 127-133, 2022.

[7] Wang Huashu, Wang Shaoshuang. “A study on the Composition and Cultivation of Translation Technology Competence in the Information Age,” East Journal of Translation, issue 1, pp. 11-15, 2016.

[8] Wang Shaoshuang, Qin Jianghua. “Constructing Technological Competence System for Translators in the Context of Big Data --- Review of A Practical Course in Translation Technologies,” Technology Enhanced Foreign Language Education, issue 1, pp. 90-96, 2018.

[9] Qiliang Cui. “A Study on the Translators’ Technical Practical Competence in the AI Era,” Foreign Languages and Cultures, vol 4, issue 1, pp. 104-110, 2020.

[10] Vocational Education Professional Profile (2022 Revision), 2022.

[11] Feng Quangong, Liu Ming. “Towards the Construction of A Three-dimension Model of Post-editing Competence,” Foreign Language World,issue 3,pp.55-61,2018.

[12] Zhong Wenming, Shu Chao. “The Competence Structure and Curriculum of Post-Editing -- Based on the Frontier Analysis of Foreign Post-Editing Courses,” Technology Enhanced Foreign Language Education, issue 6, pp. 68-91, 2020.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230200

Influence of the Targeted Poverty Alleviation Policy on the Livelihood of Poor Farmers in Western China

Jin'e MA and Qingfeng HOU[1]
The school of Marxism of Gansu Agricultural University, Lanzhou730070, China

Abstract. This paper surveyed 589 farmer families from three different areas in Western China to analyze the influence of the targeted poverty alleviation policy on livelihoods of low-income families. We can conclude that the policy has led to an obvious increase in the total income of low-income families, as well as positive effects on those poor families who rely on agriculture and government subsidies. Yet the policy showed little impact on low-income families that rely on non-farm income. The program has also encouraged farm families to increase their various sources of capital stock. Hence, we suggest that the Chinese government institute policies that provide more balanced growth and support a more sustainable development path for the rural economy and society. This might involve increasing human capital, increasing rural employment, and promoting a more efficient allocation of rural resources.

Keywords. Targeted poverty alleviation, livelihood, influence

1. Introduction

The Chinese government has been focusing on the alleviation of poverty in fourteen extremely poor areas for a long time. Three of these areas are the Qinba Mountain Area, Liupan Mountain Area and Tibetan Area, which span many provinces in north-west China. Poverty is widespread in these areas and there are many economic challenges. Since the implementation of the targeted poverty alleviation policy in 2014, the life pattern of farmers and herdsmen in these areas has changed markedly. An important part of the program is the use of residence officers who live in the rural areas and research ways of improving livelihoods. They develop improvement plans that involve distribution of farm inputs, improvements in technology, and even distribution of clothing and food. As a result of these efforts, rural livelihoods have been improved. but it is not clear that these investments have generated sustainable increases in the economic status of poor farmers. This is a significant issue not only for researchers but also the Chinese government. The long-term, sustainable effects of these policies is vitally important to future poverty alleviation schemes.

Chinese scholars have documented and explained the poverty problem [1] (Zhou 2002); focused on specific anti-poverty policies and actions [2-5] (Hong 2003, Zhu 2004,

[1] Qingfeng Hou, the school of Marxism of Gansu Agricultural University, Lanzhou 730070, China; 373418622@qq.com.

Du 2005, Xiang 2013); researched ways to measure and construct the poverty index [6] (Lu 2007); and assessed the achievements of the poverty alleviation programs [7-11] (Wei et. al. 2017, La et. al. 2018, Ni et. al. 2018, Hu et. al. 2017, Zhang et. al. 2017). But research that analyzes the effect of targeted poverty alleviation interventions on the livelihood of low-income families over time and space remains relatively unexplored. This research examined 38 villages from the Qinba Mountain, Liupan Mountain, and Qinghai Tibetan Areas and acquired detailed farmer data from interviews in 2016 and 2017. We use the difference in difference (DID) model to analyze the economic changes between targeted and non-targeted families before and after the targeted poverty alleviation policy. The goal of this research is to make a valid assessment of the effects of the targeted poverty alleviation policy in these areas. The research will also explore sustainable development methods for low-income families to see whether China's overall development goals are being met.

2. Literature Review

Analyses of targeted poverty alleviation policies mainly focus on the objectives of the policy and an assessment system for measuring results. The World Bank (1996), United Nations (1997), Asian Development Bank, Organization for Economic Cooperation and Development, Overseas Development Agency, and German Technical Assistance Company have all made great efforts in assessing poverty policies. Based on Chinese national data, Yang (2011), Zhang and Wang (2013), Hu and Fan (2014), Bai and Li (2013), Wang (2015) explored and developed an effect assessment system for targeted poverty alleviation based on Chinese national data [12-16]. Bai and Li (2013), Wang (2015), Liu and Chen (2016), Qing et. al. (2010), Ma (2002) did similar work based on the regional data [15-18].

Santa Clara City, California, brought about the comprehensive performance model (CPM) of assessment for poverty alleviation in 1998. This model includes three aspects: defining the mission, establishing goals, developing performance measures [19] (DJ Bernstein,2000). Ravallion (2008) used the propensity score matching (PSM) method to assess poverty alleviation efforts of developing countries, which emphasized the importance of methods and data collection for these assessments [20]. Habibov (2009) used the data envelop analysis (DEA) method to measure the influence of poverty reduction projects on poverty incidence [21]. Park (2009) used the matching method and panel data to analyze the influence of public financial poverty alleviation projects on low income and rural families, and concluded that these projects had not increased the income of these very poor people [22].

The assessment systems for targeted poverty alleviation developed by the academic world are comprehensive but too complicated and it lacks systematic valuation. This paper uses the relatively comprehensive and systematic sustainable livelihood theory established by Chambers and Conway in the 1980s. The Department for International Development (DFID) developed the sustainable livelihood frame in 1990 and Scoones perfected this framework to be more organized and simplified. This research will bring together Chinese and international theory and assessment methods using the difference-in-difference (DID) model to provide a more accurate analysis of targeted poverty alleviation policies.

Among all the theories of livelihood, the sustainable livelihood framework designed by DFID is relatively comprehensive by dividing livelihood assets into five aspects including nature, material, financial, human, and social capital. Natural capital means the land, water and biological resources that can be used to maintain human livelihoods, including renewable and non-renewable resources. Material capital means assets created through the productive progress of humans, including housing, infrastructure, means of production and equipment. Financial capital is the most convertible form of capital including cash, cash equivalents and credit capacity. Human capital is the knowledge, skills, labor capacity and physical condition that people reach through various investments and strategies. Social capital is the process system that is formed through social networks. Poverty and vulnerability are created by imbalances within the five livelihood capitals. Therefore, by internal development and external intervention, the livelihood capital structure of farmers can be perfected so that sustainable development is attained by disadvantaged groups.

Key findings from Chinese research on livelihood change is that internal and external factors must change to improve farmer well-being. Li, Li and Zuo (2004) formulated the livelihood paths of farmers and point out that there are many ways for improvement – expanding agricultural operations to utilize available family labor, exiting farming and moving out of the community, cooperating with other farms in concentrating agricultural production to attain scale, and diversifying their employment beyond agriculture. Qin and Mao (2009) agree that there are various ways to promote livelihood change for Chinese minorities [23]. Meng (2008) suggest that factors including ecology, technology, history and culture can influence the livelihood change of the Shui nationality [24]. Tian (2011) believes that the livelihood pattern of the Tu nationality changed based on the adaptation of local ecological environment to the social environment of Tibet and China [25]. Liu (2012) discovered that climate change had influenced the livelihood of farmers and herdsmen according to on-site research of stock farming and agricultural production in Erdos [26].

3. Data Sources, Research Technique and Variable Selection

3.1. Data Sources

Data used in this research come from surveys (taken in 2016 and 2020) of 589 farmers in Qinba Mountain Area and Liupan Mountain Area in Gansu Province and Qinghai Tibetan Area. The farmers were chosen randomly from a geographically stratified sample of the areas. The research area includes 6 counties and 29 villages. In the sample are 372 targeted families and 217 non-targeted families. We acquired the criteria for the targeted poverty alleviation program and basic information about the villagers from the data archives, but the data used in the analysis come from the questionnaires and in-depth interviews.

Qinba Mountain Area and Liupan Mountain Area encompass subtropical humid, subtropical semi-arid, subtropical monsoon and mainland monsoon climates. The mean annual precipitation is 358mm and the mean by villages varies between 335 and 450 mm. The sample villages are mainly in mountainous regions and hilly lands in remote locations. The climate in Qinghai Tibetan Area encompasses inland semi-arid and plateau continental climates. The average elevation of sample villages is over 2860m and the annual average temperature is 3.27℃, varying between -1.2℃ and 7.8℃. The mean annual precipitation is 358mm and the village average varies from 335mm to 450mm. The villagers in all three areas live in very dispersed areas and public infrastructure lags other areas in China. Communications is limited between these villages and the outside world, and economical and social development lags. During the survey, we also interviewed village heads, model farmers and other village leaders to collect data that are hard to acquire from questionnaires, such as basic village and infrastructure conditions. Furthermore, we used these sources to cross validate the data collected from the surveys.

3.2. Sample Description

The data for the research come from the questionnaires collected from the 29 villages of this project. During the field research, the project team gave out 600 questionnaires to farmers and the number of valid questionnaires is 589, an effective rate of 98.17%. The questionnaires included information on whether it was a targeted poverty alleviation family, its location, and gender, age and education level of the family members. The statistics of the sample include: 372 targeted poverty alleviation families and 217 non-targeted poverty alleviation families. The age of interviewees was: 2.2% less than 29 years old; 11.2% aged between 30 and 39; 30.7% between 40 and 49; 27.7% aged between 50 and 59; and 28.1% aged 60 and over. For the education level: 16.5% of the interviewees had no elementary school diploma; 41.8% had an elementary school diploma; 31.3% had a junior high school diploma; and 10.2% had a senior high school or higher diploma. We also checked whether the data collected from the questionnaires had good intrinsic consistency. The result showed that the Cronbach's Alpha and the KMO value are all higher than 0.8, which means that the sample data have good intrinsic consistency.

3.3. Research Methods

This paper uses the difference in difference (DID) model to analyze the influence of the targeted poverty alleviation policy on farmer income. By comparing changes before and after the implementation of the policy for households covered by the policy and those not covered by the policy, the result of targeted poverty alleviation can be assessed. The basic idea is to divide the samples into two groups, the first group is the treatment group, which is influenced by the policy, while the second group is the control group, which is not affected by the policy. The DID estimator is the difference in change in certain factors, such as income, between the treatment group and the control group.

Let group A stand for the treatment group, and group B stand for the control group. The dummy variable, D, identifies the control group, so when D=1, it means the household participated in the poverty alleviation program, and when D=0, it means they did not participate. The dummy variable, T, identifies the time period, so when T=0, it is before the policy implementation, and when T=1, it is after the policy implementation. Thus, the regression equation for analyzing the influence of the targeted poverty alleviation policy is:

$$Y = \beta_0 + \beta_1 T + \beta_2 D + \beta_3 TD + \varepsilon \tag{1}$$

TD is the product of the two dummy variables and ε is the disturbance term, which is assumed to be randomly distributed with mean zero and constant variance. For farmers in the control group, when D=0, the model can be expressed as $\mathrm{Y} = \beta_0 + \beta_1 T + \varepsilon$, therefore, the income of farmers in the control group before and after the implementation of the policy can be expressed as:

$$Y = \begin{cases} \beta_0, if \quad T = 0 \\ \beta_0 + \beta_1, if \quad T = 1 \end{cases}$$

The average change of income for the control group from the targeted poverty alleviation policy is:

$$diff1 = (\beta_0 + \beta_1) - \beta_0 = \beta_1$$

For the treatment group, when D=1, the model can be expressed as $\mathrm{Y} = \beta_0 + \beta_1 T + \beta_2 + \beta_3 T + \varepsilon$, therefore, the income of farmers in the treatment group before and after the implementation of the policy can be expressed as:

$$Y = \begin{cases} \beta_0 + \beta_2, \quad if \ T = 0 \\ \beta_0 + \beta_1 + \beta_2 + \beta_3, \quad if \ T = 1 \end{cases}$$

The average change of income for the treatment group from the targeted poverty alleviation policy is:

$$diff2 = (\beta_0 + \beta_1 + \beta_2 + \beta_3) - (\beta_0 + \beta_2) = \beta_1 + \beta_3$$

An important assumption of the DID model is that the without policy effect, the treatment group and control group should have the same average effect. Hence, the net effect of the targeted poverty alleviation policy on low income families can be shown as the TD parameter, β_3, the DID estimated value. It measures the effect after the implementation of the policy, also known as the net policy effect.

The fixed effect model is used in order to control for other factors during the process of empirical analysis:

$$Y_{it} = \beta_0 + \beta_1 T_t + \beta_2 D_i + \beta_3 T_t D_i + \beta_4 X_{it} + \varepsilon_{it} \tag{2}$$

In the equation above, i stands for farmers and t stands for time. The income of farmer i in time t can be expressed by Y_{it}, the group dummy variable D_i was used to measure whether the farmers joined the policy. When D_i=1, it means that farmers participated in the policy, when D_i=0, it means the farmer did not. The time dummy variable T_t stands for the period before or after the implementation of the policy. When $T_t = 1$, it means after the policy and when $T_t = 0$, it means before the policy. The $T_t D_i$ coefficient is the DID estimate of the policy's effect on farmer income. X_{it} is a set of control variables that influences income change.

3.4. Variable Selection

Dependent variable. The main object of the research is to explain farmers livelihood. The livelihood of farmers in the study area come from agricultural livelihood, non-agricultural livelihood, and government-supported livelihood that relies on subsidies. Total family income of farmers was used to measure the alleviation effect; agricultural income was used to measure agricultural livelihood; non-farm income was used to measure non-agricultural livelihood; and policy income was used to measure government-supported livelihood.

Control variables. Livelihood theory suggests that the main factors that affect farmer livelihood assets, which includes natural, material, financial, human, and social capital. Measures of natural capital include per capita family land area, the condition of the trees and grassland, and the subjective condition of the natural environment. Material capital includes value of fixed assets, the living space, and housing structure. Human capital includes the household's labor force, education level, and health condition. Social capital includes the neighborhood's subjective condition, farmers' participation in the village committee, and the infrastructure's condition. Definitions for all the explained and control variables are given in Table 1.

Table 1. Definition of major variables in DID model.

Name	Measure	Explanation
Explained Variables	Effect of targeted poverty alleviation policy (Y)	Total family income
	Agricultural livelihood (Y1)	Agricultural income
	Non-agricultural livelihood (Y2)	Non-farm income
	Government-supported livelihood (Y3)	Subsidy income
Explaining variable	Treated	Experimental group =1, control group=0
	Time	Before experiment=0, after experiment=1
Control variables	Family arable land per capita (X1)	Agricultural acreage/family population
	Condition of trees and grasslands (X2)	More=0.9, no change=0.6, less=0.3
	Condition of natural environment (X3)	Excellent=1, better=0.75, general=0.5, poor=0.25
	Family fixed assets value (X4)	Observed value
	Living space (X5)	Observed value
	Housing structure (X6)	Brick-concrete=1, post and panel=0.75, civil structure=0.5, makeshift house=0.25
	Family cash holdings (X7)	Observed value
	Farmers debt and credit availability (X8)	Have ability=0.9, hard=0.6, no ability=0.3
	Total family labor force (X9)	Observed value
	Householder education level (X10)	University=4, senior high=3, junior high=2, elementary=1, no school=0
	Health condition (X11)	Good=1, chronic=0.75, other=0.5, disabled=0.25
	Neighborhood condition (X12)	Good=0.9, common=0.6, bad=0.3
	Participation in village committees (X13)	Often=1, occasionally=0.75, seldom=0.5, never=0.25
	Infrastructure condition (X14)	better=1, no change=0

3.5. Descriptive Analysis of Variables

Table 2 provides the survey results by province, Qinghai and Gansu. The income figures show that there is a great difference between the farmers surveyed in the two provinces. Total household income for the Qinghai respondents was 58,572.52 yuan before the poverty alleviation program and 78,534,54 yuan after the program (an 8.52% increase). Over 76.58% of the total household income for the Qinghai respondents was from agriculture. Total household income for the Gansu respondents was 26,536.62 yuan before the poverty alleviation program and 38,104.36 yuan after the program (a 10.9% increase). Over 87.48% of the income from these farmers comes from non-farm activities, their agricultural income did not increase as much as their total household income.

Total household income for total respondents was 41,257.32 before the poverty alleviation program and 50,532.2 after the program (a 5.63% increase). Over 53.48% of the income from these farmers comes from non-farm activities. The income-earners in these households are mostly migrant workers, though they do have farm plots and herd livestock.

Table 2. Survey results by province.

area	period	Per capita disposable income (¥)	Household agricultural income (¥)	Household non-farm income (¥)	Family policy income (¥)	Total household income (¥)
Qinghai	Before experiment	15797.57	36346.67	13568.51	8815.32	58572.52
	After experiment	19888.00	44308.56	21567.98	12658.00	78534.54
Gansu	Before experiment	6508.00	4407.42	24979.23	1375.59	26536.62
	After experiment	9341.00	4785.18	31699.65	1619.54	38104.36
total	Before experiment	9993.73	16392.61	20697.37	4167.35	41257.32
	After experiment	13299.05	20582.10	24188.38	5761.71	50532.20

Table 3 shows summary statistics of the entire sample. Household income for the sampled farmers increased from 45,028.35 yuan to 64,608.55 yuan from 2016 to 2020. So, household incomes were 19,580.2 yuan higher, a 43.48% increase. Agricultural income increased from 15,808.7 yuan to 30,055.48 yuan; non-agricultural income increased from 25,521.4 yuan to 30,854.82 yuan; and policy income increased from 3698.25 yuan to 3698.25 yuan.

Table 3. Index and the descriptive statistics.

Variables Name	Measurement Index	Before the Policy		After the Policy	
		Mean Value	Standard Deviation	Mean Value	Standard Deviation
Explained Value (Y)	Total Family Income (Y)	41257.32	3327.25	50532.20	5356.95
	Agricultural Income (Y1)	16392.61	3200.4	20582.10	5366.29
	Non-agricultural Income (Y2)	20697.37	1255.07	24188.38	1259.6
	Policy Income (Y3)	4167.35	173.31	5761.71	173.31
Explaining variable (DID)	**Treated**	**Treated =260, Contral =300**			
	Time	**Before the Policy=2016, After the Policy=2020**			
Control Variables (X)	Family arable land per capita (X1)	24.25	60.32	19.74	70.31
	Condition of the of trees and grasslands (X2)	0.52	0.14	0.93	0.192
	Condition of natural environment (X3)	0.61	0.09	0.86	0.324
	Family fixed assets value (X4)	25312.59	41700.96	27810.84	52445.8
	Living space(X5)	79.54	38.06	100.56	42.34
	Housing structure(X6)	0.47	0.345	0.78	0.26
	Family cash holdings (X7)	22831.78	41456.7	26119.46	42558.29
	Farmers debt and available credit (X8)	0.44	0.14	0.75	0.26
	Total family labor force(X9)	3.03	1.44	3.04	1.44
	Householder education level(X10)	0.31	0.15	0.42	2.28
	Health condition(X11)	0.90	0.23	0.95	0.26
	Neighborhood condition(X12)	0.87	0.20	0.88	0.22
	Participation in village committees(X13)	0.57	0.11	0.86	0.23
	Infrastructure condition (X14)	0.35	0.12	0.99	0.08

The amount of farmers' livelihood capital also increased with the condition of the trees and grassland in natural capital increasing from 0.52 to 0.93, and the subjective condition of the natural environment increasing from 0.61 to 0.86. The per capita land area of households fell from 24.25 to 19.74, which is likely due to a government policy to move some cropland into forests, which would decrease land area for the household. The value of material capital increased from RMB 25,312.59 to RMB 278,10.84, and the housing area increased from 79.54 square meters to 100.56 square meters. The amount of household cash in financial capital increased from 22,831.78 yuan to 26,119.46 yuan, and the household borrowing ability increased from 0.44 to 0.75, indicating that the borrowing conditions for the farm households improved. The average educational level of household heads increased from 0.31 to 0.42, and the health status increased from 0.9 to 0.95. The neighborhood conditions changed little between the two years, but the condition of infrastructure and participation in village committee meetings changed significantly. These measures of social capital show a significant improvement over the two years.

4. Model Results and Analysis

4.1. Basic Estimation Result of DID

The DID model presented in equation (1) was fitted using the Stata15 quantitative analysis software. The estimate of the difference in difference is $ß_3$ in equation (1); the estimate of the average effect for the treatment group (targeted farmers) is $ß_2$ in equation (1); and the estimate of the average time effect is $ß_1$ in equation (1). The results of the estimation are shown in Table 4. The columns show the results for the four dependent variables, Y, Y1, Y2 and Y3.

Table 4. Results of the basic DID model.

	Y	Y1	Y2	Y3
DID	0.0142* (0.0167)	0.0134** (0.0092)	-0.0179 **(0.0086)	0.0314***(0.0112)
Treated	0.0036 (0.0028)	0.0034***(0.0013)	-0.0038 ***(0.0016)	0.0051***(0.0016)
T	0.0121 (0.0108)	-0.0059***(0.0031)	0.0119 (0.0065)	-0.0039 (0.0042)
cons	-0.0051**(0.0018)	-0.0018***(0.0009)	0.0013 (0.0015)	-0.0046***(0.0008)

***, ** and * stands for p<0.01, p<0.05 and p<0.1, respectively. The numbers in parentheses are standard errors.

The DID estimate for the effect of the targeted poverty alleviation on total household income is 0.0142and the coefficient is significantly different from zero at the 10% level. This means that the average household receiving the targeted poverty assistance had 1.4% more income in 2020 than those households outside the program. Thus, the program was successful in increasing household incomes.

The poverty alleviation program increased agricultural income by 0.0134 and the coefficient is significantly different from zero at the 5% level. So, the average household receiving the targeted poverty assistance had 1.3% more income in 2020 than those households outside the program. This percentage is slightly lower than the overall increase in household income from the program.

The program reduced non-agricultural incomes by 0.0179 and the coefficient was significantly different from zero at the 5% level. Thus, targeted households on average reduced their non-farm income by a relatively large amount (1.8%). The biggest increase in income for targeted households was in the form of government subsidies. The coefficient for government subsidies was 0.0314 and the coefficient was significantly different from zero at the 1% level. So targeted households received 3.1% more subsides than households outside the program.

When comparing the targeted poor families with non-poor families, the targeted poverty alleviation policy has obvious positive effects on agricultural livelihood and government-supported livelihood of poor families. After participating, total family income, agricultural income and policy income of poor families increased faster than non-poor families. While the effect on non-agricultural income was obviously negative, which means that those participating in the policy likely reduced that work efforts off the farm relative to non-poverty families.

4.2. The Estimated Result When Bring in the Control Variables

The results of the DID model in equation (2) is shown in Table 5. The independent variables in the model are the family's arable land per capita (X1), condition of the trees and grasslands (X2), condition of the natural environment (X3), fixed assets value (X4), living space (X5), housing structure (X6), family cash holdings (X7), farmer debt and available credit (X8), family labor force (X9), householder education level (X10), family member health condition (X11), neighborhood condition (X12), participation in village committees (X13) and infrastructure condition (X14).

Table 5. Results of DID model with control variables.

	Y	Y1	Y2	Y3
DID	0.0385***(0.0113)	0.0184***(0.0049)	-0.0112**(0.0061)	0.0323***(0.0089)
Treated	0.0085***(0.0018)	0.0062***(0.0021)	-0.0021 (0.0023)	0.0053***(0.0024)
T	0.1342**(0.0631)	-0.0231 (0.0242)	0.0119 (0.0213)	0.1512**(0.0645)
X1	-0.0353 (0.0801)	-0.0645**(0.0285)	-0.0889 (0.0621)	0.1221* (0.0136)
X2	-0.0216*(0.0123)	0.01779*(0.0089)	-0.0007 (0.0079)	-0.0368***(0.004)
X3	0.0115 (0.0162)	0.0046 (0.0082)	0.0072 (0.0079)	0.0015 (0.0169)
X4	0.0638 (0.0698)	0.1023**(0.0554)	-0.1131*(0.0692)	0.0415 (0.0336)
X5	0.0341 (0.0471)	-0.0376*(0.0213)	0.0741 (0.0499)	-0.0028 (0.0291)
X6	-0.1236***(0.0457)	-0.0204 (0.0182)	-0.0671 (0.0546)	-0.0371(0.02791)
X7	1.329***(0.2168)	0.7539***(0.1022)	0.3781* (0.2742)	0.2115***(0.0817)
X8	0.0343 (0.0384)	-0.0103 (0.0138)	0.0271 (0.0252)	0.0157 (0.0369)
X9	0.1295***(0.0506)	0.0654***(0.0236)	0.1382***(0.0321)	-0.0671 (0.0412)
X10	-0.0721***(0.0216)	-0.0141 (0.0092)	0.0251* (0.0153)	-0.0827***(0.0207)
X11	-0.1002**(0.0489)	-0.0016 (0.0121)	0.0211**(0.0099)	-0.1259**(0.0551)
X12	-0.0203	0.0007	-0.0128	-0.0092

	Y	Y1	Y2	Y3
	(0.0342)	(0.0182)	(0.0096)	(0.0328)
X13	0.0091	0.0032	0.0026	0.0041
	(0.0258)	(0.0082)	(0.0099)	(0.0282)
X14	0.0311**(0.0153)	0.0206	0.0031	0.0123
		(0.0172)	(0.0113)	(0.0135)
Cons	0.0249*	-0.0043	0.0009	0.0304* (0.0158)
	(0.0144)	(0.0045)	(0.0042)	
R- squared	0.4877	0.6011	0.1751	0.1601

***, ** and * stands for p<0.01, p<0.05 and p<0.1, respectively. The numbers in parentheses are standard errors.

The results for the estimation of model (2) shows that the estimated effect of the targeted poverty alleviation policy on total livelihood (Y), agricultural livelihood (Y1), non-agricultural livelihood (Y2), and government-supported livelihood (Y3) are identical in sign and similar in magnitude to the results without the control variables. The elasticities for total, agricultural, non-agricultural, and government-supported livelihood are 0.0385, 0.0184, -0.0112 and 0.0323, respectively. The coefficients are generally more significant with the control variables, where all of them are significant at the 5% level and three are significant at the 1% level. Again, when comparing poor families with non-poverty families, the targeted poverty alleviation policy has obvious positive effects on the total family income, agricultural income and policy income. The effect on non-agricultural income stays negative, so the policy reduces average non-agricultural income of poor families.

The results for the control variables show that higher family cash holdings, more family laborers and better infrastructure condition have a positive and significant influence on the household income. The housing structure, household head education level, and the family health condition have a negative and significant influence on household income.

Many of the control variables significantly influenced agricultural income. The condition of the trees and grasslands, fixed assets value, and family labor force number had positive and significant effects on agricultural livelihood. The family's per capita land holdings and the size of their living space had significant negative effects on agricultural income. Households with more natural capital, material capital and human capital have more income.

Family cash holdings, family labor force, education level of the household head and the family's health condition have positive and significant impacts on non-agricultural livelihood. The family's fixed assets were the only control factor which negatively influenced non-agricultural livelihood. Thus, financial and human capital have a major influence non-agricultural income of farmers.

The family's arable land per capita has a positive and significant effect on policy income while the household head's education level and the family's health condition have significant negative effects on policy income.

4.3. Robustness Test

In order to analyze the stability or robustness of the regression results, we replaced total family income with agricultural income and gradually increased the number of control variables to test the stability of the DID results. We began with natural capital controls (E), and progressively added material capital (M), financial capital (F), human capital

(H) and social capital (S). Table 6 shows that the DID regression results changed very little. This suggests that the regression results are stable.

Table 6. Robustness test of the model.

		Uncontrolled variable			**Gradually introducing E, M, F, H and S**		
DID		0.0128**	0.0129**	0.0168***	0.0177***	0.0191***	0.0179***
		(0.0056)	(0.0055)	(0.0054)	(0.0053)	(0.0053)	(0.0053)
E	X1		-0.0331**	-0.0476***	-0.0762***	-0.0721***	-0.0726***
			(0.0162)	(0.0178)	(0.0149)	(0.0159)	(0.0172)
	X2		0.0212***	0.0211***	0.0189***	0.0181***	0.0191***
			(0.0057)	(0.0057)	(0.0057)	(0.0057)	(0.0057)
	X3		0.0020	0.0028	0.0017	0.0013	0.0011
			(0.0071)	(0.0072)	(0.0071)	(0.8721)	(0.0071)
M	X4			0.1186***	0.1026***	0.1125***	0.1031***
				(0.0189)	(0.0189)	(0.0193)	(0.0194)
	X5			-0.0109	-0.0058	-0.0078	-0.0072
				(0.0125)	(0.0135)	(0.0124)	(0.0135)
	X6			-0.0462***	-0.0463***	-0.0462***	-.0462***
				(0.0123)	(0.0121)	(0.0113)	(0.0114)
F	X7				0.1398***	0.1415***	0.1408***
					(0.0222)	(0.0221)	(0.0221)
	X8				-0.0175*	-0.0155	-0.0152
					(0.0096)	(0.0096)	(0.0097)
H	X9					0.0454***	0.0453***
						(0.0159)	(0.0161)
	X10					-0.0073	-0.0072
						(0.0199)	(0.0199)
	X11					-0.0048	-0.0051
						(0.0131)	(0.0131)
S	X12						-0.0025
							(0.0112)
	X13						0.0061
							(0.0106)
	X14						0.0281
							(0.0631)

***, ** and * stands for $p<0.01$, $p<0.05$ and $p<0.1$, respectively. The numbers in parentheses are standard errors.

5. Conclusion and Advice

5.1. Conclusion

It is found by on-the-spot investigation that the poverty-stricken peasants' overall income has been increased dramatically due to the government's targeted poverty alleviation policy, especially those who rely heavily on agricultural production as their main way of living and benefit from government subsidies. But the income increase is not so obvious for those peasants whose way of living is not dependent on agricultural production. Therefore, the income of peasantry living in Zang (Tibetan area of Qinhai Province) has not increased as much as those living in Qinba and Liu Panshan Mountainous areas. As a result, the targeted poverty alleviation policy exerts a stronger impact on the peasants and herdsmen in Qinhai Tibetan District than peasantry living in Qinba and Liu Panshan Mountainous areas in Gansu Province.

Different livelihood assets have different influences on livelihood of poor peasantry. Because the time interval is relatively short before and after the assessment, the increase in human capital is very low. However, the increase of human capital has a greater impact on the income of poor peasantry compared with off-farm livelihoods. Meanwhile, financial capital plays an important role in increasing income of peasants with agricultural livelihood than those with non-agricultural livelihood. Despite the great changes before and after assessment of material capital, natural capital and social capital, through regression analysis, found that they have little effect on increasing income of poor peasantry.

Through a comprehensive analysis, the research points out that the change of livelihood assets is dramatic since the implement of the targeted poverty alleviation policy on sample peasant households while the financial capital and material capital have been improved a lot. The natural capital and social capital obviously changed to a certain degree as well. The change of human capital is relatively stable due to the limitation of the time interval before and after the program. Because human capital plays a key role in increasing income of poverty-stricken peasantry, giving priority to improving the human capital of poor households and promoting balanced development of the five types of livelihood capital are major strategies for constructing sustainable livelihoods of poor peasantry. In this analysis of the targeted poverty alleviation policy on peasantry's livelihood, the conclusion is not final due to the following factors: firstly, the time interval for the panel data is relatively short; and secondly, small sample size, which is also the key point of future research.

5.2. Suggestion

It is suggested that improvements in rural human capital in poor areas is to be focused on equalization of public service. Secondly, in order to uproot poverty thoroughly and solve the problem of intergenerational transmission of poverty, the necessary material and fund support should be provided for those students from poverty-stricken families who enter cities to pursue better education as an escape from substandard basic education in poor areas. Thirdly, to add cultural activities to peasants' life during their slack farming season to improve peasants' cultural quality. Fourthly, strengthening training of laborers so that the members from poor families have more skills.

Furthermore, the following measures should also be taken. One, to accelerate rural land system reform by advancing market allocation of agriculture production factors, clarifying the ownership of the land, breaking the dependence between peasants and their land; guaranteeing the production elements can be circulated not only in the countryside but also between countries and cities. Putting the sparsely-distributed pieces of land together thus the natural capital of peasants engaging in planting will be increased.

The peasants' enthusiasm, creativity and internal growth momentum of the economy will be stimulated by encouraging peasants to participate in some activities such as public service. One of the most effective measures is to diversify peasants' cultural activities, improve their cohesion and excite their own power of developing. Besides, to solve the problems of peasants' relatively inactive participation in public affairs resulting from unbalanced allocation of resources, the widening gap over incomes year by year and the lack of the sense of identity among neighbors' and so on. To promote management level over democracy, solve rural social justice and equality. As a result, a stable social order and a harmonious life in the countryside can be achieved and social capital can also be promoted.

References

[1] Zhou Y. Poverty research: The conflict between structural explanation and cultural explanation. Sociological Studies. 2002 Mar; (3):49-63.
[2] Hong DY. The poverty alleviation in urban China since 1978. Sociological Studies. 2003 Jan;(1):71-86.
[3] Zhu XY. New strategy of poverty alleviation: From “reducing massive poverty" to targeting the poorest of poor. Sociological Studies. 2004 Mar;(2):98-102.
[4] Du Y, Cai F. The transition of the stages of poverty reduction in rural China. China Rural Survey. 2005 Jul;(5):2-9.
[5] Xiang LL, Deng X. Anti- poverty dynamic assessment of western minority areas. Guizhou Etmnic Studies. 2013 Jan;(1):98-102.
[6] Lu KQ. Poverty indices: Construction and re-construction. Sociological Studies.2007 May; 22(4): 1-20.
[7] Wei MX, Li MW, Yang MZ. Index system construction and practice of precise poverty alleviation in Hebei province based on performance evaluation. Journal of Anhui Agricultural Science.2017 Dec; 45(24):242-245.
[8] La JJ, Que GX, Li Q. Preliminary research on performance evaluation system for the targeted poverty reduction in west China. West Forum on Economy and Management. 2018 Jan;(01):15-22.
[9] Ni XF, Kong XL. Fuzzy evaluation on the performance of poverty alleviation in poverty-stricken areas: A case study of 12 counties in Anhui province. Rural Economy and Science Technology. 2018 Jun;29(5):157-159.
[10] Hu SP, Hang L. Research on the construction of performance evaluation index system of targeted poverty alleviation in China. Economic Research Journal. 2017 May; (4): 82-90.
[11] Zhang Z. Study on the performance evaluation of targeted poverty alleviation in villages and towns of Guizhou province: Based on the data of Nangong town, Taijiang county. Economic & Trade. 2017 Jan;(1):111.
[12] Yang Y. Public expenditure, economic growthand poverty-empirical study on the 2002-2008 data of poverty-counties in China. Journal of Guizhou University of Finance and Economics. 2011 Jan;(1):88-94.
[13] Zhang WB, Wang SG. Poverty alleviation policy, income distribution and poverty reduction in rural China. Issues in Agricultural Economy. 2013 Mar;(2):66-75.
[14] Hu XY, Fan YZ. Empirical analysis on China’s rural poverty alleviation fund use efficiency. Journal of Central South University of Forestry Technology (Social Sciences). 2014 Apr; 8(3):76-80.
[15] Bai ZZ, Li CY. Research on the performance evaluation of the government's poverty alleviation funds in the new stage of poverty alleviation and development: A case on Sangzhi county, Hunan province. Minzu Tribune. 2013 Jun;(5):89-93.
[16] Wang JP. The effect of reducing poverty of government poverty alleviation funds in Tibetan area of the northwest Sichuan province as an example. Decision-Making & Consultancy. 2015;(2):40-48.
[17] Liu L, Chen ZC. Use of the poverty relief funds and its effect on poverty reduction: A case study of the rural areas in Xinjiang province. Research of Agricultural Modernization. 2016 Jan;37(1):17-22.
[18] Ma XH. Refliction on development and funds' utilization of aid the poor programme in Ethnic areas. Guizhou Etmnic Studies. 2002 Feb;(2):45-48.
[19] Bernstein DJ. Local government performance measurement use: Assessing system quality and effects: [dissertation]. Washington, DC: School of Business and Public Management, George Washington University, 2000 Mar.
[20] Ravallion M. and Chen S. China’s (Uneven) progress against poverty, The World Bank, Policy Research Working Paper NO. 3408(2004).
[21] Habibov NN, Lida F. Comparing and contrasting poverty reduction performance social welfare programs across Juris-dictions in Cannada using Data Envelopment Analysis (DEA): An exploratory study of the era of devolution. Evaluation and Program Planning, 2009 Apr;33(4):457- 467.
[22] Park A, Wang SG. Community-based development and poverty alleviation: An evaluation of China’s poor village investment program. Journal of Public Economics, 2010 Oct; 94(9-10):790-799.
[23] Qin HZ, Mao SZ. The Changes of livelihood mode of ethnic minorities from reform and opening: Investigation in Namen village, Long'an county, Guangxi province. Thinking. 2009 Jan;(1):32-36.
[24] Meng AJ. Traditional way of living of shui ethnic group and its evolution. Journal of the Central University for Nationalities (Philosophy and Social Sciences Edition). 2008 Mar;35(3): 41-45.
[25] Tian JQ. The structural changes in the livelihood of the Tu nationality in Gansu province. Journal of Yunnan Normal University (Humanities and Social Sciences). 2011 Mar;43(3):69-74.
[26] Liu HM, Wang LX, Yang J, Liang CZ, Wang W. Influence of climate change on farming and grazing households and its adaptation: A case study in Uxin banner in Inner Mongolia. Resources Science. 2012 Feb;34(2):248-255.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230201

The Influence of Tax Incentive Policy on the Development of Alternative Fuel Vehicle Enterprises—Taking Chinese Companies as an Example

Jun MA [1] and Igor MAYBUROV
Department of Financial and Tax Management, Ural Federal University, Ekaterinburg, 19 Myra Street, 620002, Russia

Abstract. In the process of promoting the development of China's alternative fuel vehicle industry, the government has issued preferential tax policies that are beneficial to its industrial development, and actively guided its industrialization process. The purpose of this paper is to research and analyze the preferential tax policy of the alternative fuel vehicle industry, optimize the allocation of resources, and promote the rapid realization of energy saving, emission reduction and environmental protection. This paper adopts the case analysis method, selects X, a listed company of alternative fuel vehicles in China, as a case, and analyzes the company's development in the past ten years, including its operation, financial situation, and tax burden. The research hypothesis is that preferential tax policies have a positive effect on the development of Alternative Fuel Vehicle enterprises. Through financial indicators and tax burden, the impact of tax preferential policies on X is analyzed, and the current problems are pointed out. Research has confirmed that preferential tax policies have a certain role in promoting the profitability of enterprises, but enterprises are too dependent on policies, and there are problems such as poor policy effects and weak pertinence.

Keywords. Tax incentive policy, the alternative fuel vehicle, tax burden, profitability

1. Introduction

With the rapid development of the automobile industry, environmental and energy problems are becoming more and more serious. China's dependence on imports of oil and natural gas has increased year by year, and has risen to 73% and 43% in 2020. At the same time, automobiles are the main contributor to the total pollutant emissions [1]. The emergence of alternative fuel vehicles can make a significant contribution to solving resource and environmental protection problems, and the government has also actively launched various preferential policies to support it. This paper adopts the case analysis method and selects X, a listed company of alternative fuel vehicles in China, as a case to

[1] Corresponding author: Jun Ma, Department of Financial and Tax Management, Ural Federal University, Ekaterinburg, 19 Myra Street, 620002, Russia; E-mail: ma_dreama@hotmail.com.
The research funding from the Ministry of Science and Higher Education of the Russian Federation (Ural Federal University Program of Development within the Priority-2030 Program) is gratefully acknowledged.

analyze the impact of preferential tax policies on the tax burden of alternative fuel vehicle enterprises, and analyzes the problems and solutions faced by enterprises.

The purpose of this paper is to research and analyze the preferential tax policies of the alternative fuel vehicle industry, on the basis of theoretical and practical analysis, to propose tax preferential policies that are in line with China's actual conditions, optimize the allocation of resources in the development of the alternative fuel vehicle industry, and promote the gradual formation of Industrial chain, and promote the rapid realization of energy saving, emission reduction and environmental protection.

Based on a microscopic perspective, this paper selects X as a representative case to more intuitively analyze the impact of tax policies on the development of alternative fuel vehicle enterprises. At the same time, this paper combines the latest tax policy, observes the effect of policy implementation in time, and provides a basis for the improvement of subsequent policies.

This research is conducive to the rational planning and use of national taxes, and to improve the use efficiency of national taxes, so as to better play the incentive role of policies, promote the faster and better development of China's alternative fuel vehicle industry, and enhance the competitiveness of China's international automobile industry.

Leontyeva confirmed that the tax incentive policy, as one of the effective tools for the government to carry out macro-control, plays an important role in guiding the development of the industry. From the perspective of the alternative fuel vehicle industry, as an emerging industry supported by the state, tax incentive policies can reduce corporate tax costs, increase corporate disposable cash flow, and promote the development of alternative fuel vehicles [2,3].

Mayburov believes that the government has a positive role in promoting the development of alternative fuel vehicle enterprises through fiscal and taxation support. The current preferential fiscal and taxation policies are not completely reasonable. To further promote the development of alternative fuel vehicles, the government must continue to introduce more targeted and effective fiscal subsidies and preferential taxation policies, and improve the structure of fiscal and taxation policies [4-6].

Through cost-benefit analysis, Yan Shiyu [7], Urwah Khan et al. [8] concluded that tax incentives can help electric vehicles reduce production costs, increase product sales, and improve the environment at the same time.

Most scholars believe that preferential tax policies can promote the development of alternative fuel vehicles; giving alternative fuel vehicle companies and consumers tax incentives, or increasing the rate of fuel tax to increase the cost of using oil. This can not only promote the development of alternative fuel vehicles, but also play a role in energy conservation and environmental protection [9-11].

2. Methods

The research of this paper on the development of X's alternative fuel vehicle is not only theoretical research, but also should effectively apply the theoretical research results to enterprise practice, based on the in-depth discussion of practical problems, and try to solve. Based on the collection of X's financial reports from 2012 to 2021, this paper uses the external characteristics of transportation, the scientific development concept, and the idea of sustainable development as a guide to study the preferential tax policies related to alternative fuel vehicle. The research methods are as follows:

Literature research method. Analyzes and collects domestic and foreign information about alternative fuel vehicle development strategies, provides indispensable materials and theoretical basis, and these literatures play a very important role in the basic framework of this paper.

Case study method. Select an alternative fuel vehicle enterprise for case analysis, and put forward tax preferential policy suggestions to promote industrial development. Case studies provide a variety of verifiable data from direct or indirect observations of individual entities, which can be further embedded in clear messages ending with derivatives [12].

This paper analyzes from the two directions of profitability and tax burden. The calculation formula is as follows:

VAT burden rate = (VAT payable - VAT refund) / current business income

Corporate Income Tax Burden = Corporate Income Tax Fee / Total Profit × 100%

3. Results and Discussion

X Co., Ltd., mainly engaged in automobile business including alternative fuel vehicle, traditional fuel vehicle, rechargeable battery and photovoltaic business. In 2021, the revenue was approximately RMB 216,142 million, a year-on-year increase of 38.02%, of which the revenue of automobiles, automobile-related products and other products business was approximately RMB 112,489 million, a year-on-year increase of 33.93%. X has an irreplaceable representative position in China's alternative fuel vehicle market. This paper analyzes the impact of tax policies on X from the perspectives of profitability and tax burden. The paper selects the data disclosed in X's corporate financial reports from 2012 to 2021, because the alternative fuel vehicle industry in China has gradually developed since 2012. Before that, the scale was relatively small, so it is not considered. From 2012 to now, it covers the main process of the development of China's alternative fuel vehicle industry.

Profitability represents the ability of a company to make a profit. From an enterprise's point of view, the most direct purpose of enterprise development is to maximize profits while maintaining stable operation of the enterprise [13]. The ability of an enterprise to obtain income is an important criterion for measuring the soundness of an enterprise's financial status.

This paper selects four indicators: gross profit margin, return on equity, Ratio of profits to cost and expense, and Earnings per share to analyze the impact of preferential tax policies on X's profitability, so as to judge the company's core competitiveness and potential development prospects.

It can be seen from Figure 1 that between 2012 and 2021, X's gross profit margin, return on equity, profit margin on costs and expenses, and earnings per share have basically fluctuated in the same direction. In the past ten years, X's profitability has fluctuated greatly. From 2014 to 2016, it showed an upward trend. After reaching the peak in 2016, it began to decline. From 2017 to 2018, it showed a downward trend. It rebounded in 2019 and slightly decreased in 2021, but the overall profitability increased. In 2012, China's overall economic environment was poor, the economic development was slow, and the traditional automobile consumption market was not optimistic. Therefore, the sales volume of X in 2012 was small. After 2012, the government began to issue a series of policy measures to vigorously support the development of the alternative fuel vehicle industry, which made X company progress in the development, production and sales of alternative fuel vehicles. In 2015-2016, X's profitability has improved significantly, but with the tightening of tax policies in 2016, X is facing greater market risks. Although it actively deploys strategic plans, it is still affected to a certain extent. X has always attached great importance to the alternative fuel vehicle business segment and has increased investment in improving technology over the past few years, so even in the face of the pressure of the epidemic, its profitability can still be maintained at an ideal level.

The tax cost is usually included in the operating cost, and the tax burden reflects the proportion of the tax actually borne by the enterprise in the daily business activities to the operating income. The level of the tax burden will directly reflect whether the tax cost borne by the enterprise is suitable for the production and development of the enterprise. The tax burden level of the taxpayer or tax object is mainly reflected by the tax burden rate, which is also an important basis for formulating and improving tax policies [14].

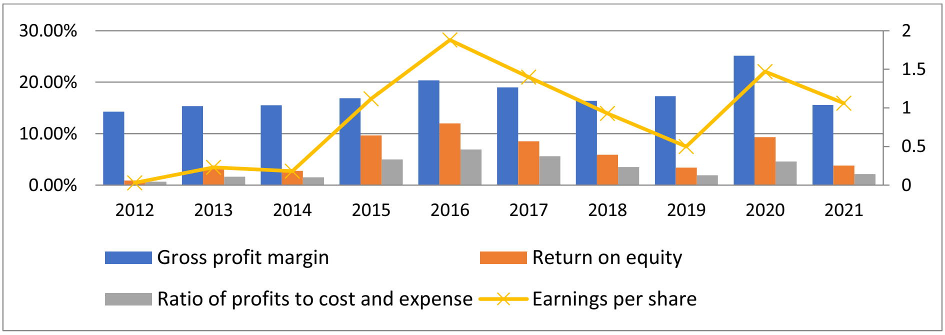


Figure 1. Changes in X's Profitability from 2012 to 2021.

Considering the availability of data, this paper selects the data in X's publicly released financial statements, and uses the effective tax rate method to assess the tax burden level of the company. Since VAT and corporate income tax account for a large proportion of China's total corporate tax burden, this paper does not consider other remaining taxes.

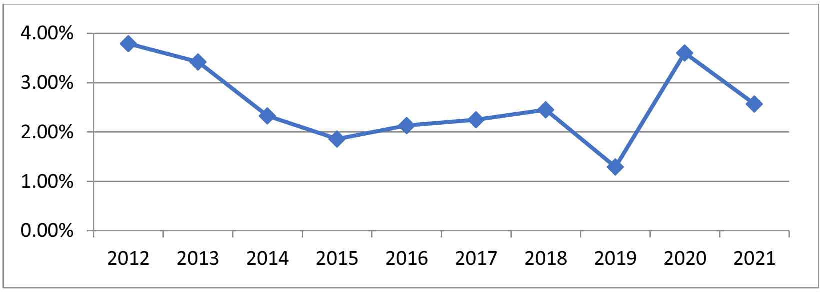


Figure 2. Changes in X’s VAT burden rate from 2012 to 2021

In Figure 2, the change of X's VAT burden rate from 2012 to 2021 shows a trend of first decreasing and then increasing. It was 3.79% in 2012, dropped to 1.86% in 2015, and the VAT burden rate increased year by year in 2016, 2017 and 2018. In 2019, affected by industry and policy changes and rising R&D expenses, it fell to 1.29%, and it rebounded in 2020. In 2121, it was reduced to 2.57% due to the increase in raw material and commodity prices. As a member of the capital-intensive and technology-intensive alternative fuel vehicle industry, X needs to pay VAT in both domestically produced and imported vehicles, and the value-added tax involves a huge amount. Therefore, the government's changes in the value-added tax policy will it has a significant impact on X's tax burden.

From 2014 to 2015, the government adopted encouraging policies to stimulate the overall development of the alternative fuel vehicle industry through policies. In 2015, X's main business income rose rapidly, and the value-added tax was reduced. Therefore, in 2015, X's VAT burden rate down from 2014. After that, X made a decision to increase the introduction of fixed assets and increase the deduction of input tax. However, the government stipulated that starting from May 2016, the real estate construction in progress of the enterprise should be deducted according to a certain proportion for two years. The reduction in the deduction is equivalent to an increase in the amount of VAT, so X's VAT burden began to rise in the three years from 2016 to 2018, and in 2018 it was even slightly higher than the level in 2014. From 2019 to 2021, China's economy is facing the triple pressure of demand contraction, supply shock, and weakening expectations, plus uncertain factors such as the spread of the epidemic, high fluctuations in international commodity prices, and geopolitical conflicts, resulting in fluctuations. Only by reasonably grasping the start and completion time of the inter-period fixed asset construction in progress can an enterprise reasonably avoid the increase of the tax amount under the background of the implementation of the policy of deducting the input tax in years.

The amount of corporate income tax and the total profit can usually be expressed as a linear relationship, and the specific situation of the linear relationship is largely affected by the income tax policy.

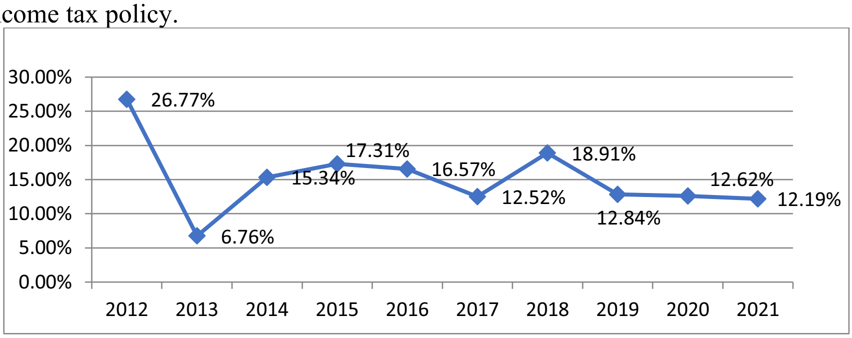


Figure 3. Changes in X’s Corporate Income Tax Burden from 2012 to 2021.

From Figure 3, it can be found that X's income tax rate in 2012 was relatively high. Due to the decline in total profits for the year, the income tax expenses continued to increase from 2013 to 2015, and the income tax expenses increased by 390% from 2014 to 2015. From 2015 to 2016, the corporate income tax expense increased by 65%, mainly due to the sharp increase in the total profit in these two years, of which the total profit in 2016, was the highest. In 2018, the corporate income tax rate increased significantly compared with 2017. In addition to the decrease in total profits, it was also because the deferred income tax expenses in 2018 were much lower than in 2017. From 2014 to 2018, X's total profit fluctuated unstable, fluctuating within a certain range, with the highest profit in 2016 and the lowest profit in 2014.

This is due to the slowdown of China's economic growth in 2014 and the great downward pressure on the economy, X adjusted the company's various businesses, and actively responded to the situation of increasing downward pressure. In 2016, the policy of tightening financial subsidies for alternative fuel vehicles was introduced for the first time, and the policy was further tightened in the following years, which had an impact on the total profit of the company. From 2019 to 2021, due to the epidemic, the income tax liability ratio is basically the same. In 2020, due to policy support, the profit is the highest at 6,882,587,000 yuan. Changes in X's corporate income tax expenses and total profits in ten years also show that the company's profits and tax burdens are affected by national policies and national macroeconomic development.

As for emerging industries, the state's macro-control directly affects their development. In order to alleviate environmental pollution, greenhouse effect, noise pollution and exhaust emissions, etc., the state has continuously issued policies to support the development of new energy vehicles.

At present, China's tax policies for the alternative fuel vehicle industry are relatively active, aiming to support the development of enterprises in the industry. However, through the case study of X, we can find some problems in the implementation of policies:

X relies too heavily on preferential fiscal and taxation policies. Once the government subsidy policy is tightened, the profit of the year will also decrease, affecting profitability; In terms of tax policy, X's alternative fuel vehicle is mostly applicable to general tax policies, which are some preferential policies for reducing or exempting vehicle and vessel tax and vehicle and vessel purchase tax. However, there are no targeted preferential policies for corporate income tax and value-added tax, which have a greater impact on enterprises. As far as tax rates are concerned, it is not found that policy changes in recent years have benefited enterprises significantly, and there is a lack of support-oriented and long-term mechanisms. The overall tax burden of alternative fuel vehicle enterprises is heavier.

Based on the specific circumstances of the case and the overall development status of alternative fuel vehicles, this paper analyzes X's own development and the tax preferential system, and puts forward specific suggestions.

As far as X itself is concerned, it insists on R&D and innovation to help companies get rid of their dependence on policies and better resist the risks of changes in external factors such as policies and markets; at the same time, expand markets and sales, improve operating cash flow; improve production efficiency, thereby reducing alternative fuel vehicle prices and increasing consumer acceptance. In terms of optimizing the preferential tax policy for the alternative fuel vehicle industry, adjust the policy focus and strengthen the support for R&D and industry infrastructure; the relatively stable implementation of Backslide Mechanism reduces the possible negative impact on the industry; accelerate the establishment and improvement of related industry supporting policies, Strive for the common development of all industries

4. Conclusion

This paper uses the case analysis method, takes X, a representative listed alternative fuel vehicle company in China as a case, combined with its data in the past ten years to study the impact of the company's current tax incentives on the company's financial indicators and tax indicators. Research has confirmed that preferential tax policies have a certain role in promoting corporate profits and reducing tax burdens, but companies are too dependent on policies, and there are problems such as poor policy implementation and poor targeting. It is recommended to strengthen the internal production and R&D of the enterprise, and adjust the focus of the policy to solve the problem. In order to alleviate energy and environmental problems by promoting the development of alternative fuel vehicle enterprises.

The promotion of the alternative fuel vehicle is still limited by many factors, such as the development level of the region and the availability of charging piles. Due to its short development time and incomplete relevant information, other relevant information will be collected in the future to conduct further research on it. This paper focuses on the analysis of the promotion effect on the alternative fuel vehicle from the perspective of tax incentive policies, and other types are less involved. Therefore, more comprehensive research will be conducted in the future.

Due to the emerging companies of alternative fuel vehicle data, the data collection is limited, resulting in insufficient research. While some results have been achieved, there is still a lot of work to be done.

References

[1] Ma J, Leontyeva YV, Domnikov AY. Analyze the impact of the transition from business tax to VAT on the tax burden of transport enterprises in various regions of China. Journal of Tax Reform. 2022;8 (2):199-211.

[2] Ma J, Leontyeva YV, Domnikov AY 2022 The Impact of China's Preferential Tax Policy on the Development of the Alternative Fuel Vehicle Industry. Journal of applied economic research, 21(2), 194-216.

[3] Leontev E, Leontyeva Y, Mayburov I. Analysis of capacity for implementation of TIF projects for electric transport development in Russian cities. International Journal of Energy Production and Management. 2021;6(4), 359-370.

[4] Mayburov I, Leontyeva Y. Fiscal instruments for regulating the sustainable development of urban transport systems in Russia. IOP Conference Series: Earth and Environmental Science. 2017;72 (012016):1-8.

[5] Ma J, Mayburov I. Using DID model to analyse the impact of vehicle purchase tax on the sales of alternative fuel vehicle. IOP Conf. Series: Earth and Environmental Science. 2021;849 (2021) 012006.

[6] Ma J, Mayburov I. Problems and prospects of China's transportation transition to sustainable development. 2021; chapter 4 p.1620-1629.

[7] Yan S. The economic and environmental impacts of tax incentives for battery electric vehicles in Europe Energy Policy. 2018; 123:53-63

[8] Khan U, Yamamoto T, Sato H. Consumer preferences for hydrogen fuel cell vehicles in Japan. Transportation Research Part D: Transport and Environment. 2020; 87, 102542.

[9] Wendler, Eugen 2014 Friedrich List (1789-1846): A visionary economist with social responsibility. Springer. pp. 12 and 135-137.

[10] Berliant M, Raa TT. A foundation of location theory: Consumer preferences and demand. Journal of Economic Theory. 1988 ;44 (2): 336.

[11] Hoffman B. Microeconomics with Calculus (2nd ed.). Addison Wesley. 1998; pp. 141-143.

[12] Ma J, Mayburov I. The impact of taxation policies on the research and development of alternative fuel vehicle companies - A case study of NIO Inc. IOP Conference Series: Earth and Environmental Science, 2022;958(1).

[13] D d'Andria, I. Savin. A win-win-win? Motivating innovation in a knowledge economy with tax incentives. Technological Forecasting and Social Change. 2018; 127: 38-56.

[14] Lapov D E, Mayburov I A. Modelling of a relative income tax bracket-based progression with the effect of a slower tax burden growth. Journal of Tax Reform, 2021;7(2), 160-172.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230202

Study on the Influence of Ash Accumulation on Radiation Quantity of Photovoltaic Modules

Lei ZHAO[a,b], Wenji CHENG[a], Qiong WU[a,1], Bo YANG[a] and Taijiang LI[a,b]
[a] *Xi'an Thermal Power Product Certification and Testing Co., Ltd, Xi'an 710089, China*
[b] *Xi'an Thermal Power Research Institute Co., Ltd, Xi'an 710054, China*

Abstract. As an important component of the photoelectric conversion process in the photovoltaic system, the working efficiency and service life of solar photovoltaic modules are closely related to the overall working performance of the photovoltaic system. Particle settlement problems caused by air pollution and abnormal weather in Northwest China increase the amount of ash on the surface area of photovoltaic modules, reducing the amount of radiation and light energy absorption. Therefore, it is necessary to carry out photovoltaic module ash accumulation experiments and radiation intensity characteristic research.

Keywords. Photovoltaic system, particle settlement, ash accumulation experiments

1. Introduction

The installation inclination, wind speed and temperature of photovoltaic (hereinafter referred to as "PV") modules are important factors affecting the surface gray of the module. Ash accumulation reduces the transmittance of the glass cover, which in turn reduces the amount of solar irradiation received by the PV cells. As one of the important parameters of PV module performance, the fodder accumulation on the surface of the module significantly reduces its efficiency and output power, resulting in greater economic losses [1-4]. Therefore, studying the change law of ash accumulation of PV panels is very important to improve their power generation efficiency.

At present, researchers at home and abroad mainly use natural fouling experiments and numerical simulation analysis methods in the research on the ash accumulation characteristics of photovoltaic modules [5-6]. In terms of natural ash accumulation experiments, Volker [7], Deline C [8] through natural ash accumulation experiments on PV modules in different regions that the loss of power generation efficiency, the decrease in light transmittance and the increase in surface temperature caused by ash accumulation have different degrees of damage and weakening effects on the daily operation and service life of the PV modules, while the ash particles adhere to the surface of the PV modules, blocking the light transmission and hindering the absorption of photons by the components is the main reason for the reduction of PV module power generation. Haeberlin [9] et al. analyzed the influence of pollutant

[1] Qiong Wu, Xi'an Thermal Power Research Institute Co., Ltd. Block A, Boyuan Science & Technology Building, No. 99, Yanxiang Road, Xi'an, Shaanxi 710054, China; E-mail: wuqiong.0702@163.com.
This paper was supported by the "Research and application of strategy to improve quality and efficiency of ash accumulation treatment in photovoltaic power plant" of independent technology projects (GR-22-TZK33).

deposition on the surface of PV modules on their power generation through natural ash accumulation experiments, and concluded that pollutant deposition will also cause certain corrosion to the glass cover plate on the surface of the module and destroy the thermal balance of the plate surface. Zhang Hao [10] built a theoretical model of the surface temperature of PV modules and the intensity of sunlight radiation, wind speed, ambient temperature and wind inclination, and concluded that the surface area ash will cause the light transmittance of the glass cover of the PV module to decrease, and the absorption of solar radiation will decrease, resulting in the rise of the surface temperature of the module. In terms of numerical simulation analysis, Meng [11] established a mechanical model of dirty particles in a dry environment, and concluded that small particles are subject to the greatest force and gravity is the second, and the two large particle size particles are subjected to the opposite force. Lu [12] used CFD simulation software to establish a dust accumulation model of PV modules, and the results showed that when the particle size of the pollution particles was less than 10μm, the amount of ash accumulation was positively correlated with the wind speed, while the large particle size was the opposite.

2. Test Construction and Instrument Selection

2.1. Test Construction

The experimental platform was built on the rooftop of a university in Northwest China. Single crystal PV modules are selected, mainly composed of PV modules, control cabinets, inverters and so on. The PV module model is: LR4-60HBD 355M, the size is 175.5cm×103.8cm×3cm, the installation angle is 10°, 20°, 30°, 60°. Since the output power of the photovoltaic module is greatly affected by the illumination intensity, ambient temperature and other factors. The weather we chose for the experiment was sunny, and the effects of light intensity and temperature were negligible. And each PV module is connected to the controller and inverter in the same way. Other factors such as line loss and mismatching can also be ignored. The module is connected to the control cabinet to adjust the electric energy generated by the PV module, and then connected to the inverter to keep the whole system relatively stable. The natural ash accumulation test bed is shown in Figure 1.

(a) 20°, 30°, 60° angle of PV module experiment bench.

(b) 10° angle of PV module experiment bench.

Figure 1. Experimental platform of photovoltaic modules from different angles.

The device connection of this experiment is shown in Figure 2.

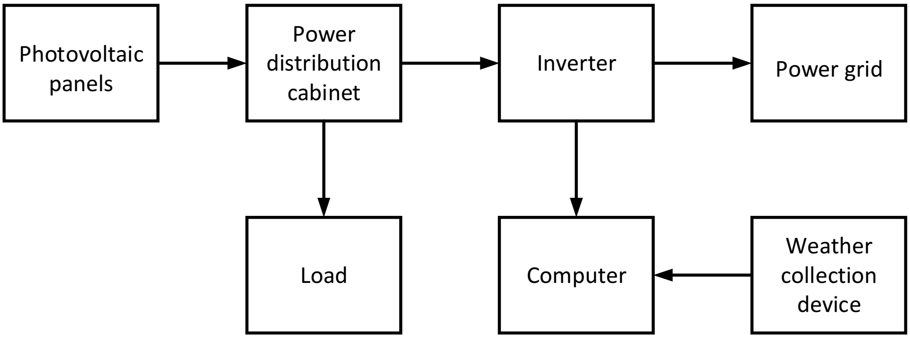


Figure 2. The device connection of this experiment.

2.2. Experimental Tools

The meteorological parameters were measured by the "yang guang" meteorological tester (TBQ-2W) and the wind speed tester (PC-4). The meteorological tester, wind speed tester, inverter collector (NBQ1X), high-precision electronic balance (BSA224S) and other experimental instruments selected in this experiment are shown in Figure 3-8.

Figure 3. High-precision electronic balance.

Figure 4. Different Angle panels.

Figure 5. Space for dust to sit.

Figure 6. Meteorological data collector.

Figure 7. Output power collector.

Figure 8. Wind speed measuring instrument.

In the process of the experiment, a large amount of sand is first piled in front of the fan with a shovel (if the sand is concentrated, the indoor dust environment can be simulated more quickly). The glass plate is placed on the smooth foam board with the surface tilted at different angles, and the fan is used to raise the fine sand on the sand bed, so as to form the sand environment. When the dust concentration in the air reaches a certain degree, the fan is turned off. After the dust settles for a certain time, the dust collecting glass plate is taken out, and the dust on each glass plate is collected and weighed by high-precision electronic balance (BSA224S). Thus, the deposition law of dust on the glass plate surface at different angles is obtained. In order to shorten the test period, we use a strong dust environment to simulate the test.

3. The Experimental Method of Ash Accumulation of PV Modules

3.1. Experimental Process and Method

Since the experiment should be conducted under the conditions of light breeze and no rain, and natural factors such as humidity should be kept stable during the experiment to reduce the error, the environment of low wind speed from August 2022 to October 2022 is selected: 1~3m/s.

Weighing the dust on the surface of the PV module: dry the electrostatic adsorption paper, measure the net weight of the electrostatic adsorption paper with an electronic scale, wipe the dust on the surface of the module with the electrostatic adsorption paper, weigh the total weight of the electrostatic adsorption paper and dust particles, and calculate the difference between the two weighing values as the net weight of dust particles.

(1) The change in the status of PV modules before and after cleaning is represented by the photoelectric conversion efficiency of PV modules. The calculation method is shown in Formula (1):

$$\eta = \eta_{ref}\left[1 - \beta_0\left(T_{PV} - T_{ref}\right) + \gamma \lg G\right] \tag{1}$$

Where, η_{ref} is the photoelectric conversion efficiency of PV module under standard test conditions; β_0 is the temperature coefficient of PV cell; γ is the solar radiation coefficient; T_{ref} is the temperature of the PV module under standard conditions; T_{PV} is the measured surface temperature of the PV module; G is the solar radiation intensity; For LR4-60HBD 355M PV module, the value of η_{ref} is 15.4%, the value of β_0 is 0.0048%/℃, the value of γ is 0.12.

(2) Experimental procedure

In order to keep the comparison conditions between the experimental results and the simulation results as consistent as possible, the dust particles on the backplane and the frame were ignored in the natural ash deposition experiment, and only the dust particles adhering to the PV module's light-absorbing surface were collected. The experimental procedure is shown in Figure 9.

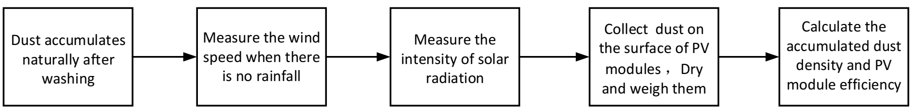


Figure 9. Schematic diagram of experimental procedure flow.

3.2. Data Collection and Collation

The density and efficiency of PV modules surface ash accumulation were selected as the evaluation indexes of the ash accumulation characteristics and the change of photovoltaic module status caused by ash accumulation. For the natural ash accumulation experiment, one week was selected as the corresponding dust accumulation time under each wind speed. The calculation results of natural ash accumulation density of photovoltaic modules at different wind speeds during the experiment are shown in Table 1.

Table 1. Natural ash density calculation results.

Wind speed	1m/s	1.5m/s	2m/s	2.5m/s	3m/s
Ash density	0.17mg/m^2	0.22mg/m^2	0.23mg/m^2	0.26mg/m^2	0.33mg/m^2

3.3. Analysis of Experimental Results of Natural Ash Accumulation

Because the wind speed will blow more dust particles to the surface of the PV module and cause deposition, the ash density obtained by the natural ash accumulation experiment increases with the increase of the wind speed. This phenomenon indicates that the change of wind speed in the environment of low wind speed has a significant impact on the surface ash accumulation density of PV modules, but the module efficiency in the environment of low wind speed is affected by the change of wind speed, compared with the change of ash accumulation density, the change range is concentrated in 0.05%~0.5%. The module efficiency change curve under the clean and ash accumulation state is shown in Figure 10.

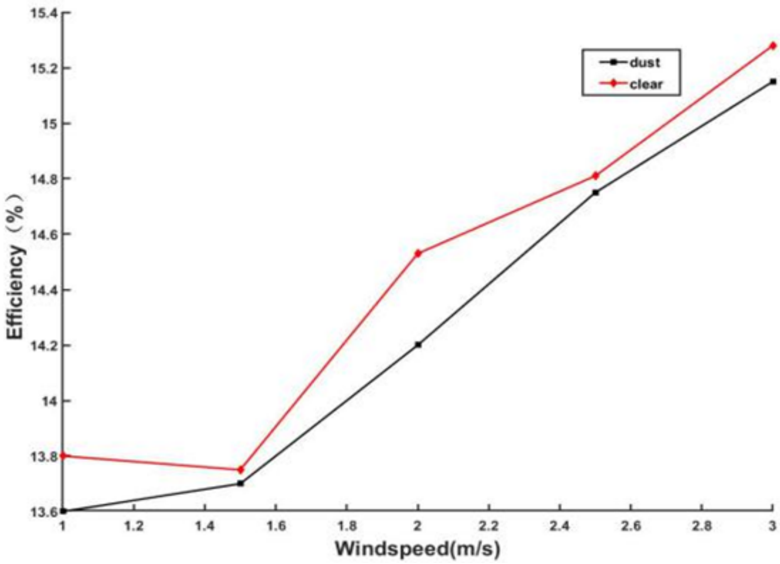


Figure 10. Schematic diagram of experimental procedure flow.

According to the analysis of Figure 10 and Equation 1, the influence of wind speed on the efficiency of PV modules is different under the two states of cleaning and ash accumulation. In the cleaning state, the increase of wind speed will reduce the surface temperature of the PV module, so that the temperature difference between the PV module surface temperature and the environment is reduced, and the efficiency of the module is improved; In the ash accumulation state, due to the shielding effect of dust particles, the thermal resistance ratio between the PV module surface and the sun increases. In addition, dust particles attached to the glass cover of the PV module surface will reduce its light transmittance. Therefore, the temperature difference between the component surface and the environment is affected by the increase of thermal resistance and the decrease of light transmittance, and the PV module efficiency decreases when the temperature difference increases. When the temperature difference decreases, the PV module efficiency increases, and finally the PV module efficiency decreases first and then increases.

4. Calculation of Radiation Received by PV Modules

The ultimate purpose of studying the characteristics of ash accumulation on the surface of PV modules is to analyze the influence of dust particles deposited on the surface of the module on its working state, and to provide reference for the method of dust removal of the module, so that it can maintain long-term economic, stable and efficient operation [13-14]. Since the working principle of PV modules is to absorb solar radiation and convert it into electric energy, which will be provided to the electricity load, the research on ash accumulation characteristics should also consider the radiation received by PV modules and the relationship between them [15-16]. In this paper, the radiation received by four types of PV modules with different installation angles on the roof of a university in Northwest China is calculated theoretically, and the variation trend and the optimal installation angle are analyzed by comparison.

About the amount of solar radiation received by PV modules, many experts and scholars have given different theoretical calculation methods. Most of the radiation data given by relevant meteorological departments are measured according to the horizontally placed modules, which is not completely consistent with the actual situation. When the photovoltaic modules are tilted, the total solar radiation on the tilted surface should be calculated H_T. By comparison, the solar radiation receiving power of PV modules with different installation angles can be analyzed. Generally, H_T cannot be directly detected by existing data, but must be measured by relevant empirical formulas or special experimental devices. In general, the total solar radiation H_T. on the inclined surface is composed of three parts, as shown in Equation 2. They are H_{bT} of direct solar radiation, H_{dT} of sky scattered radiation and H_{rT} of ground reflected radiation.

$$H_T = H_{bT} + H_{dT} + H_{rT} \tag{2}$$

The energy transfer process of the tilted PV module surface receiving solar radiation is shown in Figure 11.

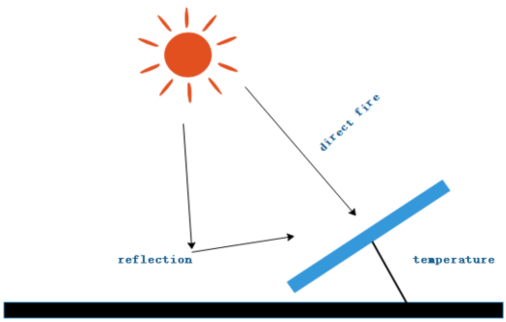


Figure 11. Diagram of photovoltaic module receiving solar radiation.

The relationship between H_{bT} and H_b of direct radiation on the horizontal plane is shown in Equation (3):

$$H_{bT} = H_b \times R_b \tag{3}$$

R_b is the tilt factor of direct radiation, which can be expressed as:

$$R_b = \frac{\left[\cos(\phi-\beta)\times\cos\delta\times\sin\omega_{sT} + \frac{\pi}{180}\omega_{sT}\times\sin(\phi-\beta)\times\sin\delta\right]}{\cos\phi\times\cos\delta\times\sin\omega_s + \frac{\pi}{180}\omega_s\times\sin\phi\times\sin\delta} \tag{4}$$

Where, ϕ represents the local latitude, β represents the installation angle of the PV module, δ represents the solar declination angle, which can be calculated from Formula 5, where n is the number of days in a year from the New Year's Day:

$$\delta = 23.45\sin\left[\frac{360}{365}(284+n)\right] \tag{5}$$

The sunset angle on the horizontal plane is expressed as:

$$\omega_s = \cos^{-1}(-\tan\phi \times \tan\delta) \tag{6}$$

The sunset angle on the inclined plane is expressed as: ω_{ST}:

$$\omega_{sT} = \min\left\{\omega_s, \cos^{-1}\left[-\tan(\phi-\beta)\right]\times\tan\delta\right\} \tag{7}$$

Sky scattered radiation H_{dT} is expressed as:

$$H_{dT} = H_d\left[\frac{H-H_d}{H_0}R_b + \frac{1}{2}(1+\cos\beta)\left(1-\frac{H-H_d}{H_0}\right)\right] \tag{8}$$

Where, H is the total radiation amount of the horizontal plane, which can be obtained by directly consulting relevant data websites, and:

$$H = H_b + H_d \tag{9}$$

In Equation 9, H_b and H_d respectively represent the amount of direct radiation on the horizontal plane. By referring to the China Meteorological Science Data Sharing Service Network, the average daily radiation received horizontally in a province in Northwest China is obtained, as shown in Table 2.

Table 2. The average daily radiation amount of a Northwest province from 2019 to 2022.

Month	Average daily radiation	Month	Average daily radiation	Month	Average daily radiation
1	5.92	5	20.71	9	15.88
2	7.84	6	17.01	10	11.16
3	12.77	7	17.44	11	8.88
4	13.54	8	17.76	12	6.78

H_0 represents the total radiation on the horizontal surface outside the atmosphere, which can be calculated as follows:

$$H_0 = \frac{24}{\pi}I_{SC}\left[1+0.033\cos\frac{365n}{365}\right]\left(\cos\phi\times\cos\delta\times\sin\omega_s + \frac{\pi}{180}\omega_s\times\sin\phi\times\sin\delta\right) \tag{10}$$

Where, I_{SC} represents the solar constant, which refers to the average distance between the sun and the earth, and the sum of solar radiant energy values of all wavelengths per unit time and per unit area. The value of the solar constant was determined by WMO as $I_{SC}=1367\pm7w/m^2$. H_{rT} represents the ground reflected radiation, as shown in Equation 11.

$$H_{rT} = \frac{1}{2}\rho H(1-\cos\beta) \tag{11}$$

Where, ρ represents ground reflectance.

The calculation of solar daily scattering is shown in Equation 12:

$$\frac{\overline{H_d}}{\overline{H}}=0.945-0.675\frac{\overline{H}}{\overline{H_0}}-0.166\left(\frac{\overline{H}}{\overline{H_0}}\right)^2-0.173\left(\frac{S}{S_0}\right)-0.079\left(\frac{S}{S_0}\right)^2 \tag{12}$$

Where, S represents sunshine duration and S_0 represents the maximum sunshine duration in a period of time. The sunshine condition of a Northwestern province was analyzed, and the total effective sunshine duration of each month was calculated and sorted out, as shown in Table 3.

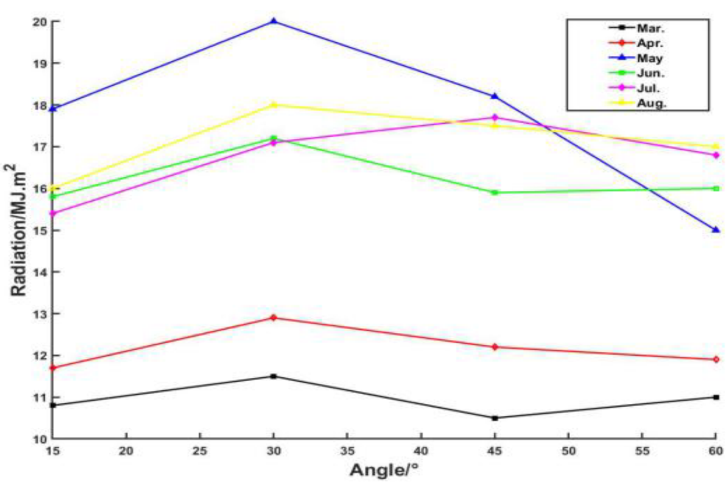
Table 3. Average daily sunshine duration from 2019 to 2022 in a Northwest province. (Unit: hour)

Month	Average monthly sunshine hours	Month	Average monthly sunshine hours	Month	Average monthly sunshine hours
1	5.48	5	8.87	9	7.56
2	6.25	6	8.51	10	7.63
3	7.11	7	7.22	11	5.31
4	7.92	8	7.13	12	4.92

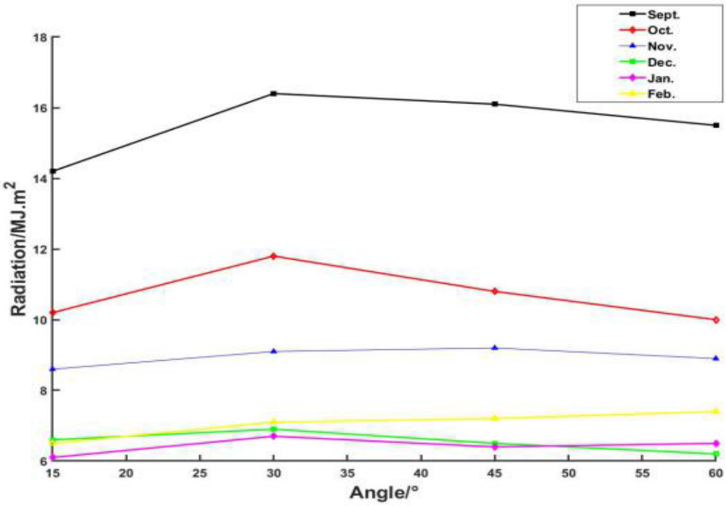
After calculating by substituting all the above formulas into relevant data, the surface radiation amount of the tilted PV module can be obtained as follows:

$$H_T=H_bR_b+H_d\left[\frac{H-H_d}{H_0}R_b+\frac{1}{2}(1+\cos\beta)\left(1-\frac{H-H_d}{H_0}\right)+\frac{1}{2}\rho H(1-\cos\beta)\right] \tag{13}$$

the final calculation results of the changes of radiation received by photovoltaic modules with different angles throughout the year is shown in Figure 12.



(a) Monthly radiation received in spring and summer



(b) Monthly radiation received in autumn and winter

Figure 12. Changes of radiation received by PV modules with different installation angles in a northwestern province throughout the year.

According to the variation trend of radiation amount data in Figure 12, the radiation amount received by PV modules with various installation angles is generally higher in spring and summer than in autumn and winter when the light is sufficient, and the optimal installation angle that makes the modules receive the highest radiation amount also changes over time. When the PV module installation angle is 30°, compared with the other three installation angles, the amount of radiation received by the inclination plane increases by 4%~12% on the whole, and it remains the largest in many months. Therefore, for a certain northwestern province, among the PV modules with four installation angles, 30° installation angle may be the best installation angle for the module to receive solar radiation.

5. Conclusion

In this paper, an experimental platform for natural ash accumulation was built, and the dust deposition law on the surface of the glass plate of photovoltaic modules under different installation angles was studied based on the experimental data. Then, the influence of ash accumulation on the power generation efficiency of photovoltaic modules under different wind speed conditions was studied, and the conclusion was drawn that the ash density of natural ash accumulation of photovoltaic modules under low wind speed environment increased with the increase of wind speed. Finally, the optimum installation dip angle corresponding to the optimum radiation quantity in a northwestern province is calculated by empirical formula. The research results of this paper provide theoretical and data basis for the subsequent simulation of ash accumulation characteristics and radiation analysis of photovoltaic modules with different installation angles.

References

[1] Lv YK, Wei ZA, Wei Z, et al. Simulation on the characteristics of natural dust accumulation of photovoltaic modules in different environments. Science Technology and Engineering. 2022;22(17):6966-6972.

[2] Ning HF, Cheng RZ, Wang WZ, et al. Experimental study on influence of dust accumulation onphotovoltaic power generation and dust removal effect. Acta Energiae Solaris Sinica. 2020;41(11):120-125.

[3] Chen JX, P GB, Ouyang J, et al. Study on prediction algorithm of dust fall on PV modules under natural rainfall. Acta Energiae Solaris Sinica. 2021;42(02):431-437.

[4] Fan SY, Wang Y, Cao SX, et al. Effect modeling and analysis of dust accumulation on output characteristics of photovoltaic modules. Chinese Journal of Scientific Instrument. 2021;42(04):83-91.

[5] Zhao WP, Lv YK, Yan WP, et al. Deposition mechanism of dust particles on the solar PV module surfaces and simulation on dust accumulation characteristics. Journal of Chinese Society of Power Engineering. 2022; 42(04): 316-325+340.

[6] Chen GS. Design of accumulated dust cleaning tester for photovoltaic modules. Technological Innovation. 2022;(10): 98-101+105.

[7] Quaschning V, Hanitsch R. Numerical simulation ofcurrent-voltage characteristic sofphoto voltaic systems with shaded solarcells. Solar Energy. 1996;56 (6).

[8] Deline C. Partially shaded operation of multi-string photovoltaic systems. Photovoltaic Specialists Conference (PVSC). 2010 35th IEEE. IEEE, 2010.

[9] Hassan AH, Rohoma UA, Elminir HK, et al. Effect of Airborne dust concentration on the performance of PV modules. Journal of Astronomical Society Egypt. 2005; 13(1): 24-38.

[10] Goverde H, Goossens D, Govaerts J, et al. Spatialand temporalanaly sisofwind effec-tson PV module temperature and performance. Sustainable Energy Technologies and Assessments. 2015; 11:36-41.

[11] Meng GS, Gao DD, Wang S, et al. Mechanics modeling of dust particle on solar panel surface in desert environment.Transactions of the Chinese Society of Agricultural Engineering. 2014; 30(16): 221-229.

[12] Lu H, Zhao WJ. CFD prediction of dustpollution and impact on an isolated ground-mountedsolar photovoltaic system. Renemable Energy. 2019; 131: 829-840.

[13] Liu WD, Luo J, Jiang XH, et al. Studying on region clustering and application of photovoltaic modules dust deposition based on its impact factors. Acta Energiae Solaris Sinica. 2022;43(01):375-383.

[14] Zhang Y, Ma ZY, Yuan CQ, et al. Reaearch progress of environmental factors on surface damage of PV modules and their protection technology. Corrosion & Protection. 2020; 41(06):7-13.

[15] Weng J, Hu Q, Wang PY, Sun DK. Design of photovoltaic module dust warning system basedon big data analysis. Science and technology innovation. 2020;(09):78-81.

[16] Wang SJ, Tian R, Guo X. et al. Dust accumulation characteristics and transmission attenuation law of photovoltaic modules. Transactions of the Chinese Society of Agricultural Engineering. 2019;35(22):242-250.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230203

Prediction of Litter Size of Sows Based on PSO Optimized RBF Neural Network

Xuzhao YANG[1]
School of Electronic Information Engineering, Xi'an Technological University, Shaanxi 710021, China

Abstract. In view of the research status of low accuracy in the prediction of litter number in pig breeding, the new method is proposed, which is based on PSO optimal RBF neural network. The litter number of sows is affected by various factors such as breed, strain, genetic gene, nutrition level, enclosure environment, feeding management level, birth age, different birthtimes, breeding management, epidemic disease and so on. According to the pig breeding information, the breeding model is established and the breeding information database is used as the input information source of neural network. In the experiment, particle swarm optimization (PSO) algorithm is used to train radial basis function (RBF) neural network. The neural network structure of 20-15-1 is established to fit the complex relationship between many indexes affecting litter size and prediction results. The experimental result shows that the error between the predicted litter number and the actual litter number is less than 5% and the prediction of litter size by PSO-RBF is in the best agreement with the actual production value. The conclusions show that the prediction method can accurately predict the litter size of sows and meet the new requirements of the development of livestock and poultry precision breeding. The research can be applied to the prediction of Litter Number in pig farm, which is helpful to improve the efficiency of pig reproduction and provide economic decision for the users.

Keywords. Pig breeding, PSO-RBF neural network, sow litter size, mean square error, precision breeding

1. Introduction

The pig breeding industry is one of the important supporting industries for the state to implement a series of policies to benefit farmers and improve agriculture, such as rural revitalization and targeted poverty alleviation [1]. China is the world's largest pig breeding country, among which the sow reserve and the pig output have exceeded half of the world's total [2]. Sows will develop healthy and healthily in the optimal growth environment, which is beneficial to improving the body performance; conversely, under excessive temperature, biting rack, catching up, slip, whipping, scare and other on [3-8]. The litter size of pig sows effectively characterized the reproductive performance and productivity of pigs, which are closely related to the economic benefits of breeding farms. The key to improve the productivity level of pig breeding is to steadily improve the reproductive performance of sows, and expecting the price increase of pigs is only expedient [9]. Sow litter performance is the key factor that determines the production

[1] Xuzhao Yang, School of Electronic Information Engineering, Xi'an University of Technology, Shaanxi 710021, China; E-mail: 414792038@qq.com.

efficiency and economic benefits of breeding farms and has a huge impact on pig quality breeding work [10].

The total litter size of reserve sows is one of the key indicators to measure the overall performance of the pig farm, and the delivery rate of reserve sows is regarded as a prediction indicator of the future performance of the pig farm. This is a good indicator to predict the future performance of pig farms. The performance of the total litter size of the reserve sows is also a good indicator, but it is lower than the delivery rate. The artificial intelligence neural network algorithm at home and abroad theory has matured, and more universities and research institutions will neural network technology is widely used in the field of livestock and poultry breeding, but few neural network algorithms applied to livestock and poultry litter number analysis, at the same time it is also artificial intelligence neural network algorithm in livestock and poultry production prediction blank [11-17]. In recent years, some large-scale livestock and poultry farms in China have gradually introduced BP neural network or RBF technology to predict the yield number of livestock and poultry, but these traditional neural network algorithms have problems such as falling into local optimal solution and local minimum value, slow convergence speed, or even unable convergence [18-25]. Therefore, the livestock and poultry breeding industry urgently need to use an objective and practical test method, namely the Radical Basic Function (RBF) based on Partical Swarm Optimization (PSO), to solve the above problems, so as to accurately predict the litter size of sows, and also to provide a basis for scientific evaluation of sow reproductive performance.

2. Materials and Methods

2.1. Sow Litter Size Influences Factor Selection

The number of livestock and poultry litter yield in rural areas generally mainly depends on the accumulated experience of breeding personnel all the year round, with strong subjectivity and low prediction accuracy [26]. Sow litter size will be pig breeding varieties, strains, genetic genes, nutrition level, season, housing environment climate, feeding management (feed type, breeding before short-term optimal feeding, pregnancy feeding amount), sow age, parity, with boar management (breeding, delivery), disease factors (disease, vaccine), herd rate, embryo survival rate of multiple factors influence and restrict [27-28]. The correlation degree between the above various elements is complex and non-linear, difficult to establish the model, and the existing evaluation methods are highly subjective and random. In order to accurately predict the litter size of live pigs, there should be a comprehensive and reasonable litter size prediction model of strong sows.

Based on early visit a large-scale pig farms, research a city animal husbandry and veterinary departments effective MoPai work, according to the influence of the existing livestock and poultry farms sow litter size factors are divided into primary influence factor and secondary influence factor, as shown in table 1, the secondary influence factor respectively using X1, X2, X3 (and so on) X20 representation.

Table 1. Influences on gilts litter size.

Impact factors	
Level 1 impact factors	Level 2 impact factors
Breeding varieties	1. Variety (sow X1)
	2. The strain (hybrid X2)
	3. Genetic factors (genetic genes are mainly X3)
Pig age parity	1.1~2 fetuses with low reproductive performance and low litter (X4)
	2.3~5 The best reproductive performance is more litter (X5 above 7 months of age)
	3. Lower reproductive performance above 6 fetuses (fatigue X6)
Poultry house environment	Temperature, humidity, illumination, and CO2、H2S、NH3Concentration (X7)
Feed management	1. Nutrition full price feed (moldy and spoiled feed is strictly prohibited) (X8)
	2. Appropriate health additives (X9)
	3. Weaning sows receive vitamins (X10)
	4. Breeding weight, age and fat condition of gilts (X11)
	5. Short-term optimal feeding for weaned sows and short-term optimal feeding before breeding (X12)
	6. Feeding amount of sows during pregnancy (restricted feeding and fat feeding X13)
Grow boars and semen	Breeding performance (3 to 5 for adults, 2 to 3 for youth, X14)
Disease situation	1. Time of vaccination (vaccine X15 during pregnancy)
	2. Blue-ear disease, swine fever, pseudo-rabies, parvovirus, swine flu, JE, swine erysipelas, etc. (X16)
	3. Serum examination (X17)
climatic factor	Spring, summer (the temperature is too high), autumn and winter (low temperature) (X18)
Piggy delivery	Pigs breathing, prevent cold and avoid suspended death (X19)
All kinds of stress	Sow slip, scare, catch up, bite and whip (X20)
comprehensive evaluation	Comprehensive prediction of sow litter size

2.2. Methods for Sow Litter Size Prediction

The PSO particle swarm biomimetic algorithm is characterized by strong global optimal search ability and good local search ability, which originates from a stochastic search model based on group intelligence collaboration designed by humans to simulate the foraging behavior of birds. The RBF feedforward neural network relies on its adaptive ability, outstanding learning ability, and is good at processing and fitting multi-dimensional data, and can achieve the consistent convergence and consistent approximation effect of any nonlinear continuous function. However, a single RBF neural algorithm is prone to fall into the local optimal solution when predicting the yield number of livestock and poultry, which will lead to certain prediction error. Therefore, the PSO particle swarm biomimetic algorithm is used to optimize the RBF feedforward neural network to improve the convergence speed and self-learning performance, so as to accurately and quickly find the global optimal solution.

2.2.1. Partical Swarm Optimization (PSO)

Partical Swarm Optimization (PSO) is an iterative research algorithm derived from the natural bird flock foraging behavior. Assuming that birds randomly search for food in an area and have only one piece of food in this area, all the birds in the flock do not know where the food is, and also do not know how far away they are from the food. When the flock catches food, the most effective way is to find as soon as possible to the surrounding area of the bird closest to the food as soon as possible. Each bird in the search area is equivalent to a possible solution to each optimization problem in the PSO algorithm, also known as a "particle".

There are several possible solutions in the problem to be found and solved by the PSO algorithm. Each possible solution represents a particle in the algorithm, and each particle corresponds to a fitness value solved based on the fitness function solution. The main characteristics of each particle are the particle fitness value, particle velocity, and particle position, where the particle velocity determines the flight direction and distance of birds. In each new iteration, the particle will independently and dynamically determine the next capture direction and position of the foraging area according to its own unique speed and position and the flying experience of other particles, so that the particle can search for the optimal solution in several possible solutions.

Specifically, each particle is self-updated by tracking the global extreme gbest and the individual extreme pkibest. When the particle searches for the above two optimal solutions, its motion rule will follow the following formula to update its own flight speed and experience position in real time:

$$\mathrm{V}_{\mathrm{i}}^{\mathrm{k+1}} = wV_i^k + c_1 rand_1()[p_{ibest}^k - x_i^k] + c_2 rand_2()[g_{best} - x_i^k] \tag{1}$$

$$\mathrm{x}_{\mathrm{i}}^{\mathrm{k+1}} = x_i^k + v_i^{k+1} \tag{2}$$

In the formula: w is the inertial weight factor affecting the particle motion velocity; $rand_1()$ and $rand_1()$ are random constant values in the interval [0,1]; c_1 and c_2 are learning factors (also known as acceleration coefficient, non-negative constant value), causing each particle to accelerate to p_{ibest}^k and g_{best}; k is the current iteration number; respectively is the flight velocity vector and experience position vector after iteration $k+1$; p_{ibest}^k is the best particle searched after iteration k; and g_{best} is the best location for the entire particle population.

2.2.2. Radical Basic Function neural network (RBF)

RBF neural network is a three-layer feed-forward analysis network with input layer, hidden layer and output layer, which can handle multi-input nonlinear function relationship. General RBF contains n input layers, m hidden layers, and 1 output layer, which can approximate any continuous nonlinear function with any accuracy. The conversion of the input to the output of the RBF network is a nonlinear relationship, while the conversion of its hidden layer to the output layer is a linear relationship. The input layer and hidden layer of RBF are connected based on Gaussian radial basis function, while the output layer and hidden layer are connected by inertial weight factor. The k-th output layer is represented by the following equation:

$$y_k = \sum_{j=1}^{n} w_{jk} \exp(-\frac{1}{2\delta_j^2} \| x - c_j \|^2) \tag{3}$$

In the formula: w_{jk} is the weight of the j th hidden layer node to the k th output layer; the c_j is the data center of the j th node radial base Gaussian function; δ_j is the width of the j th node radial base Gaussian function in the hidden layer.

Wi(i=1~m) is the corresponding weight of each layer of network. Because RBF generally only adjusts and changes the network weight, its algorithm has a short operation period and high execution efficiency. However, if the training of Wi parameters is not sufficient and the optimization is not careful, the output accuracy of a single RBF neural network will inevitably be low. Considering that the PSO particle swarm algorithm can optimize the weight factor of the RBF neural network, the organic combination of the two is adopted to solve the above problems. The factors X1~X20 in Table 1 was selected as the input parameter (20), and the sow litter number Y was selected as the output parameter. Therefore, the basic topology of multi-input single output of RBF neural network is shown in Figure 1.

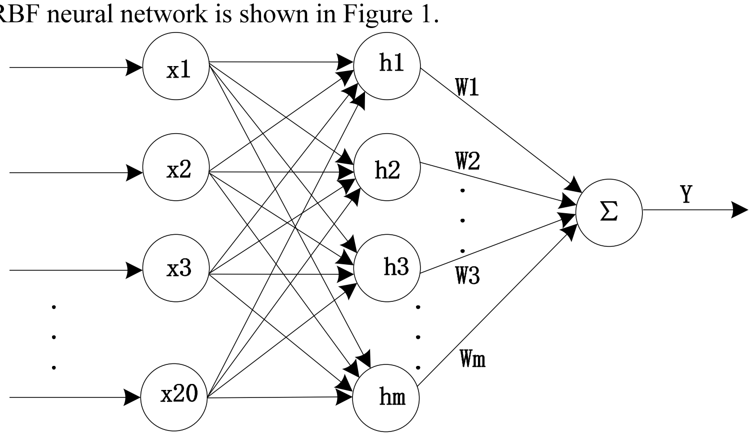


Figure 1. Basic Topology of the radial basis function RBF neural network.

2.2.3. The PSO-Optimized RBF Neural Network

The global optimization search ability of PSO algorithm is used to optimize the three important performance parameters of RBF neural network, namely, the center point cj of the radial basis function, the radial basis function width (variance), and the weight w, and determining the specific number of hidden layer nodes is also an important link in the realization of the network function. Generally, the number of nodes in the hidden layer is directly proportional to the approximation ability and output accuracy of the RBF neural network, but the more the number of nodes in the hidden layer will increase the algorithm training time, reduce the convergence ability, fault tolerance ability, generalization ability and approximation ability of the neural network, thus damaging the RBF performance. The specific number of hidden layer nodes is generally shown by the following formula. We know that the number of hidden layer nodes in the regularized PSO-RBF neural network is 6~16 (m=20, n=1 according to the actual situation of this paper).

$$k = \sqrt{m+n} + c \tag{4}$$

In the formula:c is 1 to 10; n is the output variable; m is the input variable.

The performance of the neural network is measured by training the particle fitness of the sample data. The smaller the general particle fitness value is, the closer the particle is to the best position. In the PSO-RBF optimization algorithm, its fitness function equation can be characterized by the following equation:

$$\mathrm{Fit}ness = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(train_t - y_t)^2} \tag{5}$$

In the formula: Fitness is the fitness function; n is the total sample amount of the training data; t is the t th training sample data; $train_t$ is the expected output value after the training; y_t is the actual output value.

2.3. Prediction Steps of the PSO-RBF Neural Network Algorithm

The flowchart of the sow litter size prediction by the PSO optimized RBF neural network algorithm designed according to the actual production requirements of the pig farms is shown in Figure 2.

The specific execution steps of the algorithm are as follows:

According to the practice of livestock and poultry breeding, the important influence factors X1~X15 affecting the sow litter size were selected as the input variable, and the sow litter size Y was selected as the output variable. After normalizing the training samples, the input-output non-linear correspondence of the neural network is established;

The three major parameters to be optimized in the regularized radial base neural network (center point, variance, the weights connecting the hidden layer and the output layer) are composed into each dimensional vector of the individual particle in the particle swarm optimization algorithm, and the whole PSO particle population is initialized simultaneously;

The fitness value of each particle is calculated according to Equation (6), and the Fitness value is used to determine the location of the current search area, and then the best location of the individual particle pkibest and the population particle gbest are updated based on Equation (2) and Equation (3);

According to the output value of the neural network, check whether it meets the maximum number of iterations or task completion conditions, and then the particle search process is finished and the best particle position is given in real time;

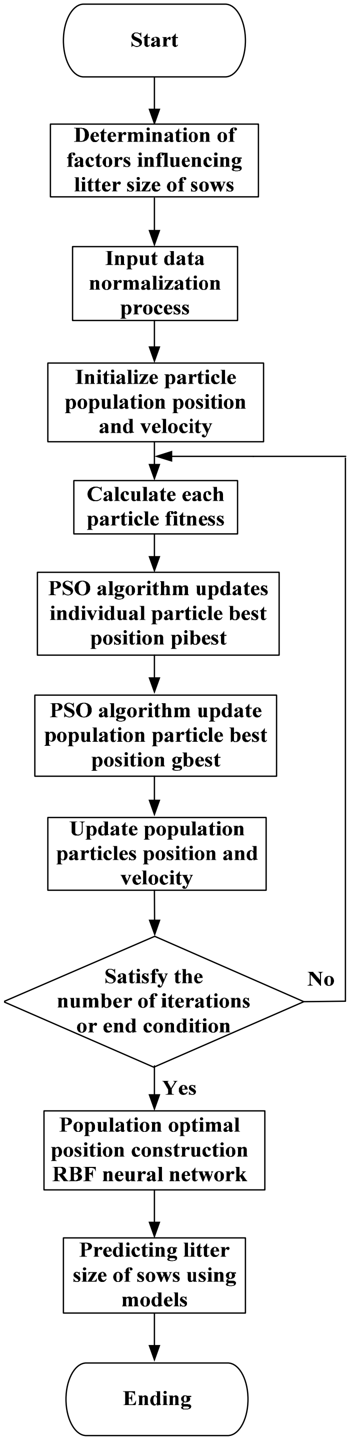


Figure 2. Flow chart of PSO-RBF neural network for predicting litter size of sows.

The best particle position obtained in (4) was taken as the optimization result to form the PSO-RBF neural network model, in which the input variable sample data was transmitted to the model to obtain the predicted value of sow litter size.

Considering that there are many input variables in this design and the inertial weight factor w will affect the motion speed and search accuracy of individual particles in the above optimization process, the convex function decreasing weight method is adopted to improve it (the particle optimization process is first fast and then slow nonlinear treatment):

$$w = (w_{\max} - w_{\min})\left(\frac{a}{a_{\max}} - 1\right)^2 + w_{\min} \tag{6}$$

In the formula: $a_{\max}$ is the maximum number of iterations; a is the current number of iterations; $W_{\max}$ is the maximum weight value; $W_{\min}$ is the minimum weight value.

3. Interpretation of Result

In order to ensure the accurate prediction characteristics of the PSO-RBF algorithm, in this paper, the number of nodes in the hidden layer from 6 to 16 are sequentially trained for the experiments, and it has been verified that different nodes and their corresponding experimental correlation coefficients R are shown in Figure 3. From Figure 3, it can be seen that the correlation coefficient value is the largest when the number of nodes in the implicit layer is 15, so node 15 is chosen as the number of nodes in the implicit layer for the training network. Therefore, a PSO-RBF neural network with a network structure of 20-15-1 is established.

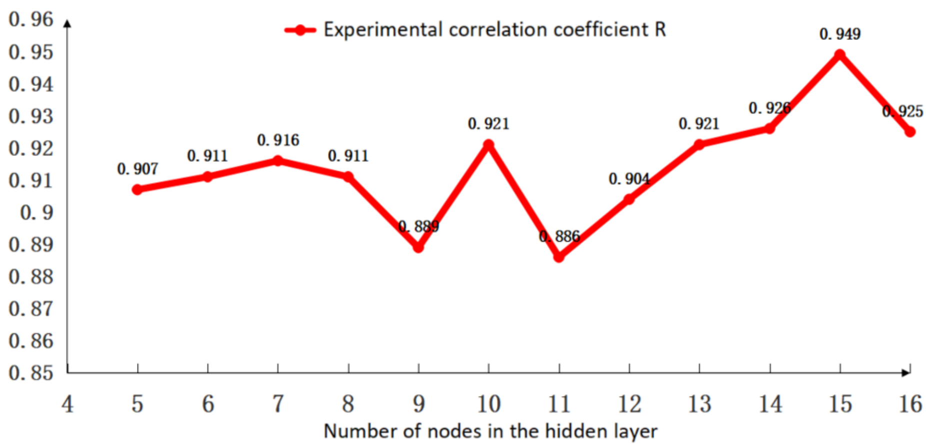


Figure 3. Different nodes and their corresponding experimental correlation coefficients.

The 20 influencing factors such as sow age and litter size, sow and gilt breed and strain, gilt semen, shed environmental information (see Table 1 for specific indexes), feed management, epidemic disease, and climatic factors were used as input variables, and only the sow environmental information could not be directly input into the PSO-

RBF neural network, while the other factors could be directly input; the output variable was one, i.e., sow litter size. According to the above analysis, the factors influencing the litter size of sows are multiple in nature. Considering that the magnitude and order of magnitude criteria are different among the influencing factors, all the above input parameters need to be normalized and pre-processed before using the neural network algorithm, and the formula is shown below:

$$P = (P_{\max} - P_{\min})Q + P_{\min} \tag{7}$$

As shown in the formula: P is the sample value; Q is the normalized value; $P_{\max}$ and $P_{\min}$ are the maximum and minimum values of the specific column where the sample is located.

3.1. Training Mean Square Error Results of Algorithms with Different Inertia Weight Control Methods

The 335 sets of breeding parameters stored in a pig farm breeding information upper computer database were used as sample data, of which 327 sets of sample data were used as learning samples and the remaining 8 sets of data were used as test evaluation samples. The prediction step using PSO-RBF neural network requires training of the neural network, and at the same time, in order to verify and compare the feasibility and effectiveness of the strategy proposed in this paper, different inertia weight factor control methods are selected to optimize the PSO network, and the specific training mean square error data are shown in Figure 4. Among the three different weight control methods, after 50 iterations, the mean square error of the algorithm based on the PSO optimized RBF neural network with inertial weight convex function decreasing strategy is less than 5%; as the number of iterations increases from 100 to 300, the mean square error of the algorithm is The mean square error of the algorithm decreases as the number of iterations increases from 100 to 300, and its prediction effect is good; it overcomes the disadvantage that the simple BP neural network algorithm is easy to fall into the local optimal solution and local minimum, and the convergence speed is slow.

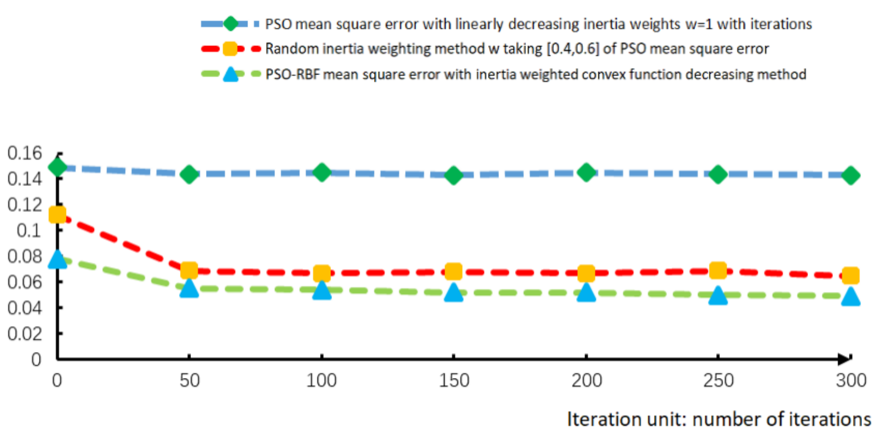


Figure 4. Mean square error corresponding to different inertia weighting factor control methods.

3.2. Analysis of PSO-RBF Neural Network for Predicting Litter Size of Sows

In order to further test the effectiveness of PSO-optimized RBF neural network, eight sets of experimental data were input into BP neural network, traditional RBF neural network and PSO-RBF neural network after in-depth training, and the prediction accuracy and working reliability of the three algorithms for litter size of sows can be compared. The distribution of prediction results and comparison of prediction curves of litter size of 64 sows in a farm by three different neural network algorithms are shown in Figure 5(a) and Figure 5(b).

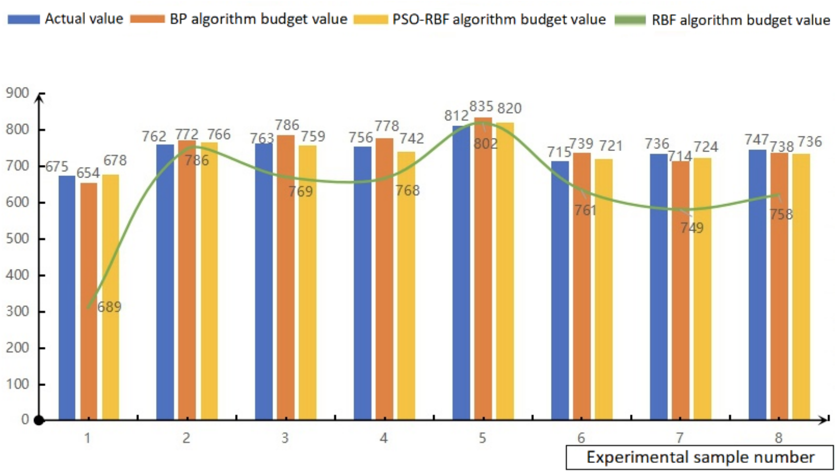


Figure 5 (a). Distribution diagram of the prediction results.

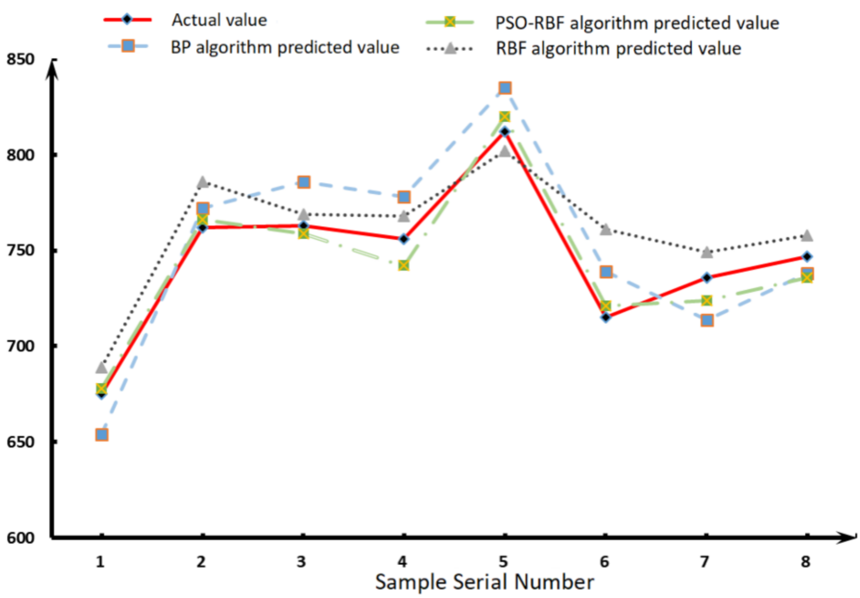


Figure 5 (b). Comparison of the prediction curves.

According to Figure 5, the prediction curve of sow litter size based on PSO optimized RBF neural network algorithm has the highest agreement with the actual situation, which also verifies the high prediction accuracy of this algorithm from the side, and has more superior convergence than the traditional pure BP neural network or RBF neural network.

4. Conclusion

In this study, the sows of a large-scale pig farm in Shaanxi Province were selected as the experimental object, which is based on the demand for accurate prediction of litter size of sow population, and 335 groups of breeding parameters were selected as the sample data. RBF neural network algorithm is used to predict the litter size of sows in this farm based on PSO optimization. When the number of hidden layer nodes is 15, the correlation coefficient value is the largest. After training with a large number of data samples, the algorithm determines the prediction neural network structure of 20-15-1 and adopts the convex function weight control strategy. Under the premise of increasing the number of iterations from 100 to 300, the prediction mean square error gradually decreases. The mean square error of PSO-RBF in predicting litter size of sows is less than 5% and the highest coincidence with the actual output value. At the same time, PSO-RBF algorithm can effectively improve its prediction accuracy and convergence speed, so this prediction method has certain practical significance and application value.

5. Future work

In the previous studies on sow breeding and production, there are few reports on how to predict the litter size of sows by establishing mathematical models. However, this study proposes an optimized neural network modeling method based on learning the traditional regression analysis method. However, due to the limited number of samples in this study, the accuracy and universality of the prediction model constructed by the new method need to be further verified; Secondly, the factors affecting litter size of sows in this study are limited, and there are few key factors that significantly affect litter size; In addition, this study is only a preliminary exploratory study using the neural network method, without in-depth parameter adjustment of the model. Therefore, the next research will focus on expanding the data volume that affects the litter size of sows, adding key factors, and gradually optimizing the prediction model. In the future sow breeding, the use of PSO-RBF neural network algorithm can provide a more objective and accurate scientific prediction of the litter size of sows in pig farms. In view of the prediction error of this neural network, the generalization ability of this measurement method should be further improved and more influencing factors should be considered in the modeling process of this study. In the future, the research focus will be on the establishment of classification standards for high and low yield sows.

6. Reference

[1] Li XJ, Wang HY, Jiang BJ, et al. Prediction of litter size traits of sows based on machine learning method. Journal of Huazhong Agricultural University. 2020;39 (04): 63-68.

[2] Han LM. Influential factors and improvement methods of litter size of sows. Modern Animal Science and Technology. 2020; 68 (08): 68-69.
[3] Zhang MF, Lu N, Sun L, He JH, Yuan XP. Adjusting sow gestation feeding levels according to body condition to maximize economic benefits of sow farms. Pig Farming Today. 2021(02):60-61.
[4] Yang Y. Factors affecting sow litter size and piglet birth weight and their improvement measures. Modern Animal Science and Technology. 2020;66 (06): 63-64.
[5] Wang XJ. Feeding management and prenatal care of gestating sows. China Animal Health. 2021; 23(03):72+75.
[6] Jiao XP. Reason analysis and improvement measures for low litter rate of sows. Modern Animal Science and Technology. 2021; 75 (03): 23-24.
[7] Tang CY, Wang JJ, Luo JR, Dan B. Key points for improving sow productivity and lifetime production performance. Foreign Animal Husbandry (Pig and Poultry). 2020; 40(12):4-11.
[8] Sun H, Song ZX, Li LH, et al. Progress report on breeding of new high-propagation large white pig line. Hubei Agricultural Science. 2019; 58 (S2): 365-368.
[9] Deng L, Zhu L, Deng Xiu D, Cao T, Ye L, Zhou JL, Zhao YX. Effects of deep sperm delivery with different densities and doses of pig semen on sow reproductive performance. Pig Science. 2021; 38(01):111-114.
[10] Xie P. Study on the effect of mixed semen of "Du+Chang" on sow litter size. China Pig Industry. 2020;15 (05): 50-52+55.
[11] Zhou KF, Hou MQ, Wang YN, et al. Investigation and analysis of the current situation of Sanyuan sow breeding and breeding. Pig Breeding. 2020; 170 (03): 79-81.
[12] Wu SF. The main measures to improve sow productivity. Contemporary Animal Husbandry. 2020(04):4-5.
[13] Han HY, Wang JJ. Otch European pig farming team. Nutritional Opportunities for Modern High Yielding Sows. Foreign Animal Husbandry (Pig and Poultry). 2020; 40(01):36-39.
[14] Wang HL, Liu XB, Chi LZ. Effect of sow nutrition on piglet weight change during pregnancy. China Feed. 2019; 640 (20): 12-16.
[15] Wei SL, Feng T, Liu Y, Tian JH. Effect of nutritional factors on litter size of sows. Pig Farming Today. 2019(04):59-61.
[16] Liu ZK. The mammary gland development problem of Sanyuan sows should be given sufficient attention. Pig Science. 2019; 36(11):126-127.
[17] Xi ZS. Introduction to key measures to improve annual productivity of sows in large-scale pig farms. Animal Husbandry Environment. 2020(04):47+60.
[18] Pedan B, Yan YQ, Cao N, Xu SY. Technical points for improving litter size in pig farms. Sichuan Animal Husbandry and Veterinary Medicine. 2019; 46(12):36+38.
[19] Kou YB. Measures to improve the fertility of sows. Sichuan Animal Husbandry and Veterinary Medicine. 2020; 47(11):40-41.
[20] Wang N, Gu L, Ming KY. Technical measures to improve the fertility of sows. Shandong Animal Husbandry and Veterinary Medicine. 2019; 40(06):25-28.
[21] Mo ZL, Huang Y, Qin XL, Hu YK, Li CB. Factors affecting the fecundity and litter size of sows and improvement measures. China Animal Health. 2021; 23(04):80-82.
[22] Song ML, Wang YQ, Song CY, Yu GH. Effect of different treatments on reproductive performance of estrous hind sows under batch production conditions. Pig Breeding. 2021(02):35-37.
[23] Ruan LD. How to improve the farrowing rate of sows in free-range mode. Animal Husbandry and Veterinary Science and Technology Information. 2020(12):141.
[24] Wang SM, Liu XH. Technical points of improving farrowing rate in large-scale pig farms. Animal Husbandry and Veterinary Science (Electronic Edition). 2020(22):28-29.
[25] Wang LW, Liu Y, Li P, Zhou G. Effects of different doses of allylgestatin on simultaneous estrus, body mass and reproductive performance of hind sows. Chinese Journal of Veterinary Medicine. 2021; 41(01):157-161.
[26] Zhao QH, Yin YS, Luo H, Lin HZ, Zhang SQ. Study on the effect of adding L-carnitine nutrient solution to diets on the reproductive performance of sows. Sichuan Animal Husbandry and Veterinary Medicine. 2021; 48(01):28-30.
[27] Li FQ. Introduction to sow breeding technology. Jilin Animal Husbandry and Veterinary Medicine. 2021; 42(02):19-20+22.
[28] Yan Y, Chen XL, Liu Y, et al. Comparative study on the litter size of three-yuan and two-yuan sows retained by enterprises under the background of African swine fever. China Pig Industry. 2021;16 (01): 46-51.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230204

Research on Storage Life Prediction of Fuze Based on Accelerated Life Tests

Junling GUO[a,1], Hongtao ZHANG[b], Yuepeng MA[c], Donghua LI[b], Weiguo YU[b] and Dongzhi ZHAO[b]
[a] *North University of China, Taiyuan 030051, China*
[b] *Hua An Industry Group Co., Ltd., Qiqihaer 161000, China*
[c] *Jin Xi Industries Group Co., Ltd., Taiyuan 030000, China*

Abstract. Fuze is an intelligent system for detection, identification and determination of targets. It is a core component of ammunition equipment and its life prediction is of great practical significance to improvement of the reliability of ammunition equipment. Based on the analysis of failure modes and failure mechanisms of fuze in natural storage environment, a storage reliability model was established for fuze. The acceleration model and the accelerated life test plan were determined. Tests were conducted in an isothermal and humidity-constant incubator to test and testing data of a fuze was processed. The results demonstrated that the storage life of this fuze was 10.05 years (prediction by Weibull distribution and maximum likelihood estimation). The proposed method improves cost effectiveness of fuze logistics and provides an important way to predict fuze life.

Keywords. Reliability, fuze, life prediction, accelerated life tests

1. Introduction

Fuze is an intelligent system for detection, identification and determination of targets. It is a core component of ammunition equipment. Fuze accelerated storage life tests aim at improving the acceleration efficiency of fuze accelerated tests, reducing the test cycle of fuze life prediction, grasping the quality change law of fuze under normal storage conditions in a short test time, and predicting storage life [1].

Through the research on accelerated storage life test of fuze technology, we obtain the reliability during fuze storage, and analyze the failure modes, failure mechanisms and failure distribution rules of equipment parts, so as to provide a basis for determining the maintenance and repair intervals of the equipment during the storage period, as well as the readiness rate of equipment and logistics support cost [2]. At the same time, it can also provide a theoretical basis for the design of a noval fuze storage reliability, maintainability and life.

When evaluating the storage life of fuze at home and abroad, the real-time 1: 1 continuous life test is mainly adopted. However, this method requires a large number of subsamples, resulting in great wastes of financial resources, material resources and time. Considering long storage period of fuze, the product is eliminated due to its backward performance before the life test is completed. Additionally, due to the rapid

[1] Junling Guo, North University of China, Taiyuan 030051, China; E-mail: 457671874@qq.com.

development of science and technology, the speed of product replacement is becoming faster and faster. Hence, people urgently need to obtain life information of products in a short time, especially the navigation system. Accelerated life test technology applies targeted stress according to the failure mechanism of the product, which greatly improves the cost effectiveness of the reliability test. Therefore, the research on life prediction has been widely valued in the field of reliability test engineering. Nelson, Meeker, et al. implemented systematic research on the optimization design of accelerated life test with the simple constant stress under the Weibull distribution and lognormal distribution, respectively, and obtained relatively satisfactory results. In terms of data processing of accelerated life test, Shi Shisong, Wang Lingling, Fei Heliang, et al. thoroughly studied various stress application methods of the accelerated life test, different life distributions obeyed by the product, and different conditions such as considering competing failures. Finally, many research results have been obtained [3]. The navigation institute of the Instrument Department in Shanghai Jiaotong University performed the life test analysis on the relatively simple dynamic tuning gyroscope, and the parameter performance extrapolation method was used to preliminary discuss the gyroscope life prediction method [4].

Accelerated life test technology is proposed to predict storage life of fuze, which greatly improves the cost effectiveness of fuze storage life prediction and has important practical significance.

2. Design of Accelerated Storage Life Test of Fuze

2.1. Selection of Stress and Method of Accelerated Life Test

The design of the accelerated life test scheme is to design an optimal test guidance scheme. Long-term practical experience indicates that when the components and raw materials of fuze are in a high temperature environment, thermal aging, oxidation, viscosity reduction, swelling, etc. will occur; low temperature increase the viscosity of the fuze material, leading to the occurrence of embrittlement or shrinkage [5]. During the temperature cycle, the expansion and contraction forces generated by the object are all concentrated on the weak link of the product, which accelerates the expansion of the weak link and causes the material failure of fuze, and affects the storage environmental stress of fuze that is mainly temperature stress [6]. Considering that fuze is fully sealed, temperature was chosen as the acceleration stress in this test.

Fuze is expensive, and has many test parameters, plenty of test times, long test cycles, as well as high precision and test fundings. Therefore, accelerated storage life test of fuze selected the step-stress accelerated life test method [7].

The fuze accelerated life test adopted the step-stress accelerated life test method with non-substitute timing truncation. The determination of the maximum stress level in weak parts of fuze can be based on the conclusion of the reliability enhancement test. The reliability enhancement test technology was mainly to take faults and failures as research objects. By applying a limit stress higher than the actual endurance to the prototype, the potential defects were stimulated in a short period of time, resulting in the weak links exposed as soon as possible, so that the working limit and damage limit of the test object were obtained, which provided the basis for determining the stress level parameters of accelerated life test in weak parts of fuze.

Fuze tactical technical indicator specified that the reliability afterstorage for 30 years is more than 0.5, and the reliability exceeds 0.85 and confidence is 0.8 with maintenance every 6 years during storage. Due to the reparability of fuze during storage, after comprehensive consideration of the storage period of 30 years and maintenance every 6 years, the total test time is determined according to 1/20 of the maintenance cycle specified by tactical technical indicator. The total test time is calculated as (test truncation time) 120 days.

The test time of the prototype under each stress level T1 is determined according to the principles of long accelerated storage life time under low stress and short accelerated storage life time under high stress. The truncation time of this test scheme under each stress level T1 is: t1 = 45d; t2 = 30d; t3 = 24d; t4 = 21d.

2.2. Determination of the Number of Sample

The step-stress accelerated life test requires only one set of samples. Therefore, the total number of samples is greatly reduced. When scheduling the step-stress accelerated life test, the number of samples cannot be too small, otherwise it will cause difficulties in data analysis.

According to the principle of accelerated life test and the selected the step test method, and considering the accuracy of data processing, the number of samples selected 40 in this test.

3. Analysis of Fault Mode and Failure

A summary of fault modes is shown in **Table 1**.

Table 1. A summary of fault modes

Number	Fault prototype number	fault modes	Reasons of faults
1	N04, N21	The working voltage and current output that provide combat fuze are both 0.	Transistor 3CG21C (01V6) 3CG130C (01V7) is open.
2	N04, N24	The working voltage and current output that provide fixed depth fuze are both 0.	Transistor 3CG21C (01V8) 3CG130C (01V9) is open.
3	N20, N34	The working current that provides microcomputer fuze is less than 45 mA.	Magnification drift of crystal triode 3CG130C (01V5)

As observed, the main factors causing faults of the accelerated life test prototypeare the open circuit failure of the transistors 3CG21C and 3CG130C, and the hFE parameter drift of the 3CG130C. The temperature stress accelerates performance degradation of physical and chemical reactions for transistors, such as ion migration, impurity diffusion, intermetallic compound synthesis, molecular changes, creep, crystalline changes, and rearrangement of microstructures in insulating materials, as well as the material aging, which results in parameter changes or failures.

4. Working Life Prediction

4.1. Calculation of Fault Occurrence Time of Fuze Prototype

The fault occurrence time of each fuze prototype is calculated as **Eq. (1)**:

$$t_{ikj} = t_{jk-1} + \frac{t_{ik} - t_{ik-1}}{r_{ik} + 1} j \tag{1}$$

where i is the sequence number of different stress levels; k is the serial number of prototype test at the same stress level; r_{ik} is the cumulative number of faults in the k-th test interval (t_{ik-1}, t_{ik}) at the i-th stress level; j is the sequence number of the fault prototype in the k-th test interval (t_{ik-1}, t_{ik}) at the i-th stress level.

The prototype fault time statistics are shown in **Table 2**.

Table 2. Prototype fault time statistics.

Stress level Ti		**T1 = 363 K**		**T2 = 368 K**	
Test number i, k	T1, 1	T1, 2	T1, 3	T2, 1	T2, 3
Test moment t_{ik}	10.06	10.16	10.22	10.29	11.12
The number of faults r_{ik}	3	1	1	1	2

4.2. Storage Life Prediction

The evaluation data of storage life of fuze was obtained entirely through the step-stress accelerated life test, and the following assumptions are made before the test processing:

- Assumption 1: Under the condition of temperature stress, the life tof the fuze prototype obeys the dual-parameter Weibull distribution, and the Weibull distribution cumulative failure function is expressed as **Eq. (2)**:

$$F(t) = 1 - e^{-\left(\frac{t}{\eta}\right)^m} \tag{2}$$

  where m is the shape parameter and η is the characteristic life.
- Assumption 2: Under different stress levels, the fuze failure mechanism is the same, that is, m in the life distribution remains unchanged under each stress.
- Assumption 3: η and T_i of the fuze are expressed as **Eq. (3)** according to Arrnenius model:

$$\eta = e^{a + \frac{b}{T_i}} \tag{3}$$

  where a is the intercept of the acceleration equation and b is the slope of the acceleration equation.

4.3. Storage Life Prediction

Since the fuze prototype adopted timing truncation tests, the storage life of all failure prototypes under accelerated stress was converted to the failure time at normal temperature. According to the numerical solution, the likelihood estimate can be obtained: a = -12.50; b = 6380; m = 3.35. Table Type Styles.

4.4. Estimation of Characteristic Life

From the acceleration equation **Eq. (4)**, the estimated characteristic life (η_0) under the normal storage temperature stress T_0 can be obtained:

$$\eta_0 = e^{a + \frac{b}{T_0}} \tag{4}$$

4.5. Working Life Predication

Define R(t) as storage reliability of fuze, then its reliability life can be calculated by **Eq. (5)**:

$$t(R) = \eta(-\ln R)^{\frac{1}{m}} \tag{5}$$

If reliability = 0.8, a storage life of fuze is 10.05 years.

5. Conclusions

Based on the analysis of the failure mode and failure mechanism of fuze in a natural storage environment, its storage reliability model is established. The storage acceleration model and accelerated life test scheme are determined, and relevant tests are conducted. Finally, based on the failure data obtained from the test and related theories, the storage life of fuze t = 10.05 is successfully predicted when the reliability is 0.8, which provides a method with important practical significance for product life prediction.

References

[1] Deng AM, Chen X, Zhang CH. A comprehensive review of accelerated degradation testing. Acta Armamentarii, 2007, 28 (8): 102-107. (in Chinese)
[2] Wei JX, Zhang J, Liu DJ. Technique analysis on reliability of liquid-floated gyro. Inertial Technology Journal, 2004, 12 (4): 57-60. (in Chinese)
[3] Zhang CH, Wen XS, Chen X. A comprehensive review of accelerated life testing. Acta Armamentarii, 2004, 25 (4): 485-490. (in Chinese)
[4] Xu JN, Bian HW, Liu Q. Principle and application of gyro. Beijing. Defense Industry Press, 2009. (in Chinese)
[5] Li M, Lin C, Li XG, et al. Infrared image on-line inspection and remaining life prediction for atmospheric heater tubes. Chinese Journal of Mechanical Engineering, 2004, 40 (12): 139-144. (in Chinese)

[6] Chen DS, Song YG, Xu Q. Gyro motor accelerated life test design based on engineering reliability analysis. Missiles and Space Vehicles, 2005 (279): 38-41. (in Chinese)
[7] Liu JJ. Researching accelerated life test of electrical connecters storage reliability. Hangzhou: Zhejiang Industry and Commerce University of Science and Technology, 2009. (in Chinese)

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230205

Analysis and Modeling of Mid- and Long-Term Output Characteristics of Photovoltaic Power Plants

Lei ZHAO[a,b], Qiong WU[b,1], Wenji CHENG[b], Bo YANG[a] and Hang XI[a]
[a] *Xi'an Thermal Power Product Certification and Testing Co., Ltd, Xi'an 710089, China*
[b] *Xi'an Thermal Power Research Institute Co., Ltd, Xi'an 710054, China*

Abstract. The study of mid- and long-term output characteristics of photovoltaic power plants is of great significance for the prediction of photovoltaic power generation and the optimal scheduling of multi-energy complementary. This paper uses the historical operation data of a photovoltaic power plant in Shandong Province to study the important factors affecting the power generation of photovoltaic power station. Furthermore, the mid- and long-term output model of photovoltaic power plants is conducted which is more suitable for engineering practice, considering temperature, dip angle and loss of each part of the power plants. MATLAB is used to simulate the model, and the results show that the output trend of the whole year can be well simulated.

Keywords. Photovoltaic power generation, historical data analysis, mid- and long-term model, engineering application

1. Introduction

As an industry that can be matured in the direction of new energy, the solar photovoltaic (hereinafter referred to as "PV") industry will further expand in today's vigorous development of energy. By the end of 2021, the installed PV power generation in China has reached 54.88GW, with a cumulative installed capacity of 305.98GW [1].

Due to the intermittent and fluctuating PV output, large-scale grid connection will bring great challenges to the smooth operation of the power grid [2-4], accompanied by the phenomenon of "abandonment of light" [5].

At present, some literature at home and abroad has studied the output characteristics of PV power plants. Tian et al. [6] analyze the output process of a single wind and PV power plant built in Qinghai Province from 2014 to 2016. According to the different time scales, the output characteristics and time complementary of wind power and PV power are described in detail. The results show that the variation of wind power generation in Qinghai Province is significantly greater than that of PV power plant, and the power generation output between wind and PV has certain complementary, which provides a technical basis for the stable operation of the power system after being connected to

---

[1] Qiong Wu, Xi'an Thermal Power Research Institute Co., Ltd. Block A, Boyuan Science & Technology Building, No. 99, Yanxiang Road, Xi'an, Shaanxi 710054, China; E-mail: wuqiong.0702@163.com.
This paper was supported by the "Research and application of strategy to improve quality and efficiency of ash accumulation treatment in photovoltaic power plant" of independent technology projects (GR-22-TZK33).

large-scale wind power and PV power. Zhao et al. [7] deals with the ambiguity of cloud cover, and a new solar radiation prediction method is proposed after the solar radiation model is modified. Xu et al. [8] analyze the output characteristics of PV output at different time scales of day, day and season based on Markov model and clustering algorithm, and a PV output clustering and simulation method is proposed. The model can well maintain the weather, timing and probability characteristics of PV output, and has higher accuracy. The research results in the above literature are mainly aimed at short-term and ultra-short-term prediction of photovoltaic. The mid- and long-term output simulation has not been studied.

Based on historical data, this paper analyzes the mid- and long-term operation characteristics of PV power plants, finds out the key factors affecting the power generation of PV power plants, and establishes a mid- and long-term output model of PV power plants that is more suitable for engineering practice considering temperature, inclination angle and loss of various parts of the power station. Finally, the simulation comparison of the model proves that the mid- and long-term output simulation proposed in this paper meets the engineering application.

2. Workflow and Characteristics Analysis

In order to explore the mid- and long-term output characteristics of PV power generation, the one-year historical operation data of a PV power station in Shandong was used for analysis. The data collection time is from January 1, 2021 to December 31, 2021, the PVoutput interval is 30min, the PV module is installed at a fixed inclination angle of 33°, the azimuth angle is 0°, and the actual installed capacity of the project is 20MW. The workflow of our entire characteristics analysis is shown in Figure 1.

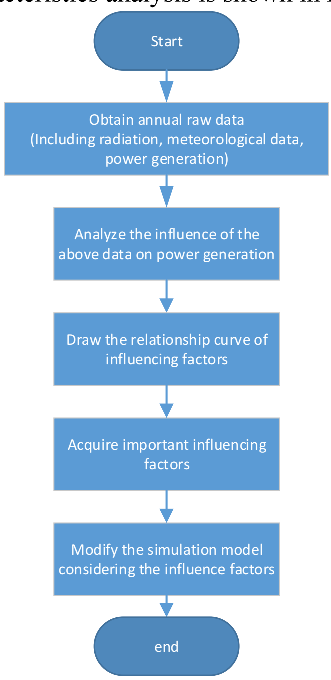


Figure 1. Workflow of characteristics analysis.

Firstly, the relationship between irradiance-temperature-power throughout the year is simulated, and the curve is shown in Figure 2. Then, the relationship between temperature and power under the same irradiance is simulated, as shown in Figure 3. Figure 4 shows the output curve of a typical day for the power plant. Figure 5 shows the three-dimensional curve of the annual power time.

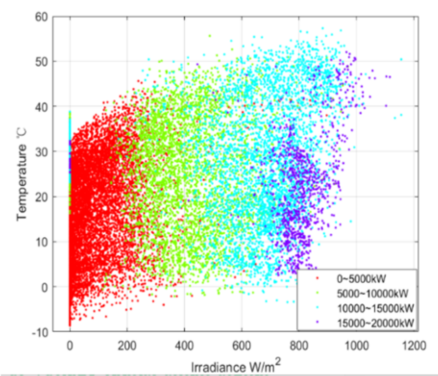


Figure 2. Irradiance - temperature - power curve.

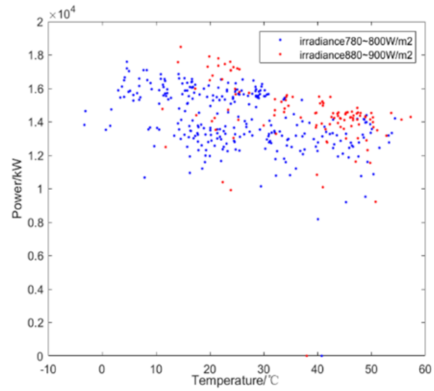


Figure 3. The relationship between temperature and power under the same irradiance.

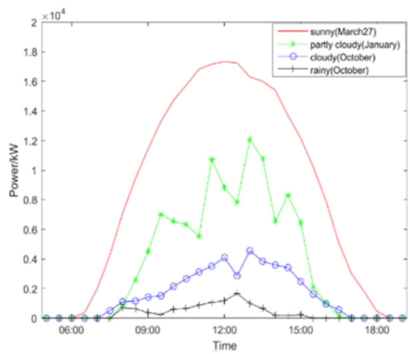


Figure 4. Typical daily output curve of photovoltaic power station.

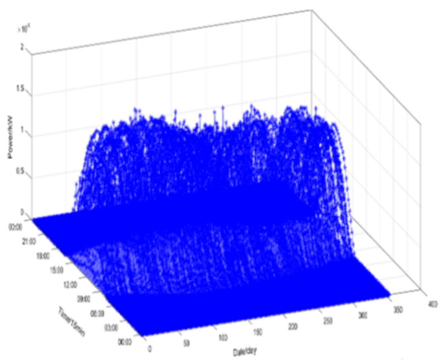


Figure 5. Three-dimensional curve of annual power time.

From the above historical data actual output curve, the following conclusions can be drawn: in Figure 2, the actual output of PV power plant is greatly affected by irradiance, and the influence relationship is almost linear, and the greater the irradiance, the greater the PV output. The temperature of PV modules also has a certain influence on the output of PV power plants, but the effect is relatively small compared with irradiance on power in Figure3. Figure 4 shows the typical weather output curve of PV power plants. The actual output curve was drawn by selecting the data of four typical days: sunny (March 27), partly cloudy (January 2), cloudy (October 16) and rainy (October 10) to draw the actual output curve. From Figure 5 it can be seen that: First, when the PV power generation output is clear and there are fewer clouds, the sunrise output curve presents a single-peak "steamed bun type", which is very smooth. The output time is concentrated at 07:00~18:00. The output reaches maximum at 12:00~13:00, and there is no sunlight radiation at night and no electrical energy output. Second, on rainy days with thick clouds, the output uncertainty is large and chaotic. Third, the summer light time in Shandong is more than in winter, and the power plant operation time is longer in summer and shorter in winter, However, due to the high summer temperature, the overall output of the maximum output period of the PV power plant is not high.

3. The Output Model of Mid- and Long-Term Photovoltaic Power Plant

According to the analysis of output characteristics, PV output is mainly related to irradiance, temperature. In order to get a more accurate output value, we also need to consider the system loss. Therefore, the influence of the above three factors must be considered when modeling PV power plants in the mid- and long-term.

3.1. Time Series Model of Irradiation

When the temperature and inclination angle of the PV panel remain unchanged, the output of the PV array is mainly related to the irradiation intensity. The irradiation intensity available for PV panels in the t period can be calculated according to the following formula [9]:

$$I_t = I_b + I_d \tag{1}$$

$$I_b = I_0 \tau_b \sin\alpha \tag{2}$$

$$I_d = I_0 \tau_{\mathrm{d}} \cos^2\left(\frac{\beta}{2}\right) \sin\alpha + \rho I_0 \tau_{\mathrm{r}} \sin^2\left(\frac{\beta}{2}\right) \tag{3}$$

$$I_0 = S_0 \left[1 + 0.033\cos\left(\frac{2\pi\left(N+10\right)}{365}\right)\right] \tag{4}$$

$$\sin\alpha = \sin\delta\sin\phi + \cos\delta\cos\phi\cos\omega \tag{5}$$

$$w = (t-12)\times 15^\circ + (120^\circ - \psi) \tag{6}$$

$$\delta \approx 23.45\sin(2\pi \times \frac{284+n}{365}) \tag{7}$$

$$\begin{aligned}\cos\theta_i = {} & \cos\beta\sin\delta\sin\phi + \cos\beta\cos\phi\cos\delta\cos\omega + \sin\beta\sin\gamma\cos\delta\sin\omega \\ & + \sin\beta\sin\phi\cos\delta\cos\omega\cos\gamma - \sin\beta\cos\gamma\sin\delta\cos\delta\end{aligned} \tag{8}$$

$$\tau_b = 0.56\times(e^{-0.56M_h'} + e^{-0.096M_h'}) \tag{9}$$

$$\tau_d = 0.271 - 0.274\tau_b \tag{10}$$

$$\tau_r = 0.271 + 0.706\tau_b \tag{11}$$

Where, I_0 indicates the intensity of solar irradiation in the outer layers of the atmosphere; S_0 is the solar constant, which represents the total amount of solar radiation received per unit area perpendicular to light entering the Earth's atmosphere, with a value of about 1367W/m^2; N is the date ordinal number of the year, and n equals 1 for January 1; α is the solar altitude angle; ω is the solar time angle; ϕ is the local latitude; δ is the declination angle; ψ is the local longitude; t is Beijing time; θ_i is the angle of incidence of the sun; β is the inclination angle of the photovoltaic panel; γ is the azimuth of the photovoltaic panel; ρ is the surface reflectivity; τ_b is the direct transparency factor; τ_d is the scattering transparency factor; τ_r is the reflection transparency factor. The above formula has taken into account the effect of inclination when modeling. If the influence of temperature, system loss and occlusion is not considered, the output of the PV power plant is defined:

$$P_{c,solar}(t) = P_{stc}\frac{I_t}{I_{stc}} \tag{12}$$

Where, P_{stc} is the output of PV panels under standard conditions, which is, the actual installed capacity of the PV power plant.

3.2. Output Model That Considers the Influence of Ambient Temperature

The effect of ambient temperature changes on the actual output is not considered in Equation 12. The ambient temperature of the power station changes in real time, so the above output expression needs to be corrected. The temperature correction output of the photovoltaic power plant at t moment can be expressed [10] as follows:

$$P_{t,solar}(t) = \begin{cases} \dfrac{P_{stc}I(t)^2}{I_{stc}R_c}[1+\partial_T(T_t - T_{stc})], & 0 \le I(t) < R_c \\ \dfrac{P_{stc}I(t)}{I_{stc}}[1+\partial_T(T_t - T_{stc})], & R_c \le I(t) \end{cases} \tag{13}$$

Where, R_c can be valued 150-200W/m^2 and T_t is the temperature of the photovoltaic panel cell at t moment; ∂_T is the temperature coefficient of the PV panel.

Notice that the temperature referred to in Equation 13 is the temperature of the PV panel cell body. The temperature read out by the meteorological information of the PV power plant is the ambient temperature, so the temperature in Formula 13 needs to be converted. A large number of experiments have proved that the function relationship between PV cell temperature and solar irradiation intensity and ambient temperature is shown in Equation 14 [11], which has sufficient accuracy from an engineering point of view:

$$T_t = T_{air} + k \cdot I_t \tag{14}$$

Where, T_{air} is the ambient temperature; k is parameter that vary according to the performance of the PV panel.

3.3. Correction Model That Considers the Losses of Each Part

Large-scale ground-mounted PV grid connected power stations are mainly composed of photovoltaic module array, intelligent confluence equipment, inverter equipment, booster stations and grid connected lines. The loss of the entire PV power plant system can be divided into four parts: PV square array loss, DC transmission loss, DC/AC inverter loss, and AC transmission loss.

(1) PV array loss

This part of the loss can be calculated by measuring the ratio of the difference between the output and input of the DC side PV array and the installed power at standard irradiance (1000W/m^2). The main reasons for this loss are as follows:

1) Module matching loss: This part is caused by the mutual blocking of PV panels, even if the shadow distance control is strictly implemented for the design and fine construction of the system.
2) Air pollutants and occlusion loss: that is the loss of dust, rain and snow on the surface of the PV modules.
3) Unusable energy loss: Solar radiation that is not available when the amount of radiation is low.
4) Temperature loss: PV modules have a negative temperature coefficient, resulting in loss due to increased temperature and reduced power.

The above four losses can be considered as losses caused by shadows, cloud occlusion, weather changes (sunny, rainy, cloudy, etc.), and dust occlusion of PV modules. We define the above losses as uncertain factors. Temperature loss have corrected in the previous section of modeling. PV system losses in this section do not take into account the losses generated by PV arrays.

(2) DC transmission loss

DC transmission loss is mainly caused by DC cables, combiner boxes, DC cabinets and other components. This loss can be measured by measuring the input and output power of each link, and then calculating the corresponding loss value.

(3) DC/AC inverter loss

The loss of the inverter mainly comes from the power loss of its internal electrical components. At present, the mainstream inverter is divided into centralized box inverter integrated machine and string inverter. By measuring the input and output power of the inverter in each power plant, the inverter loss value can be obtained.

(4) AC transmission loss

This part mainly refers to the transmission loss from the AC output of the inverter to the high-voltage grid, focusing on the loss generated by transformers and AC cables. The loss of the above links can be measured by the input and output power of each link, and then the corresponding loss value can be calculated.

According to the test results of multiple actual projects, the loss range of each part of the PV power plant can be given, see Table 1 for details.

Table 1. Loss statistics of the above parts.

Symbol	Loss type	Range of loss values (%)
μ_{dc}	DC transmission loss	9%~10.5%
μ_{inv}	DC/AC inverter loss	1%~1.5%
μ_{ac}	AC transmission loss	1.5%~2.5%
μ_z	Total Loss	11.5%~14.5%

The total loss is equal to the sum of the losses of the parts:

$$\mu_z = \mu_{dc} + \mu_{inv} + \mu_{ac} \tag{15}$$

Considering the loss of the PV power plant, the output $P_{s,solar}$ at t can be expressed as follows:

$$P_{s,solar}(t) = P_{t,solar}(t) \times (1 - \mu_z \%) \quad (16$$

4. Result and Discussion

In order to explore the applicability and correctness of the built model, the example uses a PV power plant in Shandong for verification. The specific simulation conditions are shown in Table 2.

Table 2. Parameters of PV power plant.

latitude	longitude	Installed capacity	System efficiency	Dip Angle	Azimuth Angle	I_{STC}
35.68	117.5	20MW	88%	33°	0°	1000W/m^2

The modeling curve is shown in Figure 6-8:

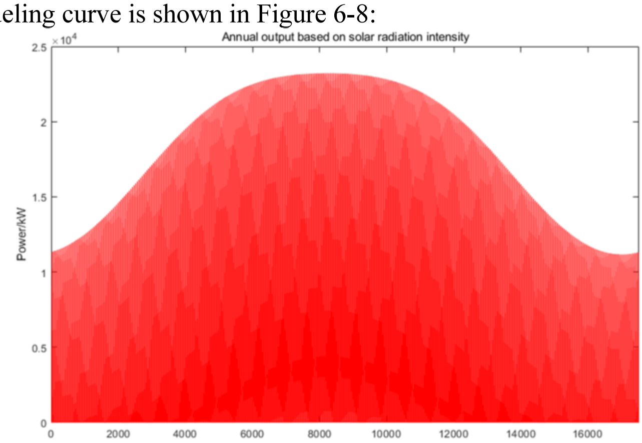


Figure 6. Annual output curve of PV power plant based on solar radiation intensity.

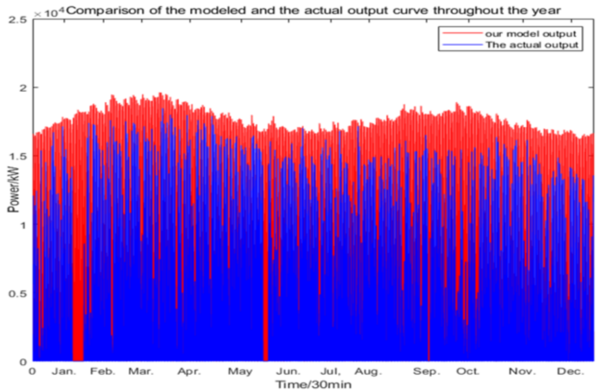


Figure 7. Annual modeling curve was compared with the actual output curve.

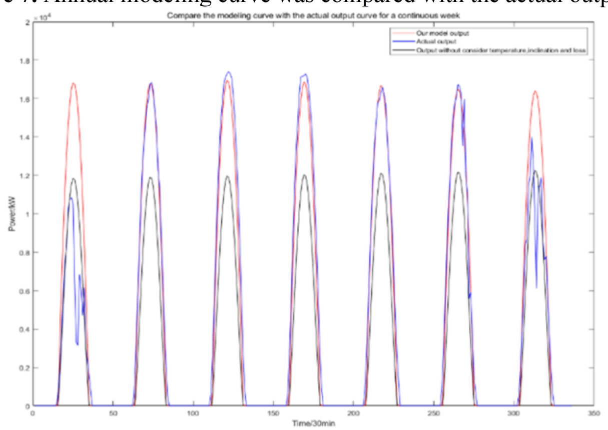


Figure 8. Compare the modeling curve with the actual output curve for a continuous week.

Figure 6 shows the annual output curve of the PV power plant based on solar radiation intensity from January to December 31, 2021 based on geographical location simulation. The curve is the maximum value that the theoretical solar radiation intensity can reach without considering temperature, inclination angle, system loss and any occlusion situation. Any PV power plants can obtain the above output outsourcing network as long as the latitude and longitude, time, installation inclination and module azimuth are given.

Figure 7 shows the annual modeling curve considering temperature and system loss, and comparing this curve with the actual annual output curve of the PV power station shows that the modeling curve is similar to the overall outline of the actual output curve, which is closer to the actual output curve than the theoretical curve in Figure 6. Here needs to be explained that because of the Spring Festival holiday the actual output curve from February 11, 2021 to February 17, 2021 did not have power generation data. From May 31, 2021 to June 3, 2021, due to the maintenance of the power plant, no power generation data was provided, so there were two obvious gaps in the actual simulation curve. However, considering that it is mid- and long-term modeling, a small number of data default will not affect the overall output trend of the power station.

Figure 8 is to compare the modeling curve with the actual output curve more clearly. The comparison chart of modeled output curve and actual output curve for 7 consecutive days from March 5 to March 11, 2021 was selected. It can be seen from the figure that the modeling curve in this paper is more accurate than the output without considering temperature, inclination and system loss. The simulation results are more accurate in sunny and cloudy days, and the simulation accuracy is 90% or more; The accuracy of cloudy weather simulation is about 80%, but in rainy days due to shadow, cloud occlusion, weather changes (sunny, rainy, cloudy, etc.), PV models dust occlusion and other factors, there are obvious random fluctuations, rainy weather simulation accuracy is low. But for mid- and long-term simulation, the main assessment of simulation trends, such simulation effect is enough to meet the guidance of mid- and long-term planning and scheduling

5. Conclusion

This paper comprehensively analyzes the important factors affecting the power generation of PV power plants through the measured data, and simulates and analyzes the typical weather and annual power generation characteristics of the PV power plants. According to the analysis results of medium and long-term output characteristics of photovoltaic power station, this paper takes into account dip Angle, temperature and various losses of the power station to model the med- and long-term output of photovoltaic power plants. By comparing with the actual power generation data, it can be seen that the modeling method described in this paper can well simulate the mid- and long-term output trend of PV power plants. The research results of this paper are of great significance for the mid- and long-term power generation prediction and multi-energy complementary optimization scheduling of PV power plants. However, in this paper, the simulation accuracy of low irradiance such as cloudy and rainy needs to be improved. Clustering and compound simulation can be used to improve the med- and long- term simulation accuracy in subsequent studies.

References

[1] National Energy Administration: 54.88GW of new PV installed capacity in 2021, with a cumulative installed capacity of 305.987GW! [EB/OL].[2022-03-10].https://guangfu.bjx.com.cn/news/20220310/1209186.shtml.

[2] Zhang Y, Chang P. Distributed model predictive load frequency control for power system with renewable energy. Electric Drive. 2022;52(20): 70-75.

[3] Li J, Ai X, Wen J, et al. Quantitative analysis of probability distribution for duration time characteristic of photovoltaic power. Automation of Electric Power Systems. 2017;41(6):30-36.

[4] Petrakopoulou F, Robinson A, Loizidou M. Simulation and analysis of a stand-alone solar-wind and pumped storage hydropower plant. Energy. 2016;96(2):676-683.

[5] Wang G, Zhang D, Li J, et al. Control strategy of photovoltaic storage power generation system considering photovoltaic power curtailment. Acta Energiae Solaris Sinica. 2019; 40(03):817-824.

[6] Tian X, Ji S, Zhang P, et al. Study on the output characteristics of photovoltaic and wind power generation in Qinghai. Northwest Hydropower. 2019; 6 (02):1-6.

[7] Zhao S, Wang M, Hu Y. Research on the prediction of PV output based on uncertainty theory. Transactions of China Electrotechnical Society. 2015; 30(16):313-220.

[8] Xu B, Zhu P, Yuan Y, et al. Clustering and simulation of photovoltaic output adopting markov model. Electric Power Construction. 2019; 40(2):54-62.

[9] Zhang X, Kang C, Zhang N, et al. Analysis of mid/long term random characteristics of photovoltaic power generation. 2014;38(06):06-13.

[10] Shen H, Zeng Z. Solar photovoltaic power generation technology. Beijing : Chemical Industry Press. 2005:34-36.

[11] Zhang X, Cao R, et al. Solar photovoltaic grid-connected power generation and inverter control. Beijing : China Machine Press. 2012:46-50.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230206

Government Thoughts on Environmental Cost Control

Lina FA[1]
School of Economics and Management, Shanghai University of Political Science and Law, Shanghai 201701, China

Abstract. The Japanese government arbitrarily releases nuclear pollution into the Pacific Ocean. It reduces the time and economic costs of enterprises, but also causes the loss of environmental costs. From the perspective of international law, this is even more illegal. This also reflects that Japanese enterprises and the government do not pay enough attention to environmental cost control. They fail to realize that environmental damage will inevitably lead to enterprises and governments needing to spend more costs to make up for it, which is not conducive to ecological environment protection and human long-term development. Therefore, from the long-term interests, the government and enterprises should bear the responsibility of controlling environmental costs. The government should play the role of macro-control and replace the end of governance with the source of governance. It plays a role in improving legal norms, strengthening supervision and guidance, establishing and improving the emission trading system, and increasing environmental information disclosure.

Keywords. Environmental cost control, marine pollution, Government guidance, Emission permit trading

1. Introduction

The Japanese government announced that the direct dumping of 1.3 million tons of nuclear waste into the sea, which violated the obligations of international law and would cause major pollution and damage to marine ecology.[1] It takes at least 20,000 years for these nuclear-contaminated wastewater to be completely diluted by marine ecosystems. This kind of behavior that will bring huge environmental costs must be paid attention to. Governments should learn from it and control environmental costs in a reasonable way, especially the control of corporate environmental costs.

Furthermore, Enterprise environmental cost control is a long-term and arduous comprehensive work that runs through each link of the enterprise. A single control method cannot achieve the development of low-carbon economy of enterprises. The government needs to play a role, combined the actual situation and specific needs of our country, from many aspects to improve the mechanism loopholes, build a reliable guarantee mechanism in line with social trends, response to the central government's call to develop circular economy, and strive to implement the international requirements of economic and social sustainable development. [2]

[1] Lina Fa, School of Economics and Management, Shanghai University of Political Science and Law, Shanghai 201701, China; E-mail: falina12@163.com.
Thank Shanghai University of Politics and Law for the convenience of finding data and literature, and thank my students for their help in finding data.

2. Cost of Marine Environmental Pollution in Japan

2.1. Increased Social Environmental Costs

In the traditional environmental cost calculation theory, the enterprise environmental cost only includes the economic cost consumed by enterprises to protect the environment. And excluding the loss of enterprise environmental cost caused by resource and environment consumption. This also makes enterprises often choose to sacrifice the environment in exchange for economic growth. This way may save the cost of enterprises in a short time, the consequences of resource depletion and environmental damage will eventually shift to the enterprise. Enterprises need to consume more cost to control environmental pollution. The choice of Japan' s nuclear power plants to discharge nuclear wastewater directly into the Pacific is closely related to cost choices. Since the nuclear leakage accident, the Fukushima Daiichi nuclear power plant in Japan has not been able to create value for Japan' s economy in its development. [3] Instead, it has cost the Japanese government nearly USD 190 billion in the past 10 years. Therefore, in order to save time and economic costs to the greatest extent, the Japanese government abandoned the traditional treatment methods of nuclear pollution such as stratum injection, drainage into the sea, steam release, hydrogen release and underground burial. Instead, Japan chose to dump nuclear wastewater into the sea and transfer environmental costs to society.

Moreover, the reality proves that the discharged nuclear wastewater cannot completely eliminate radioactive elements even after treatment. Tokyo Electric Power Company used to remove radioactive elements by setting up accident radioactive sewage purification device, but the effect is not obvious, tritium, strontium, cesium, iodine, and other radioactive nuclides still exist. These radioactive materials will also spread around the world with the water cycle.

GEOMAR, a German marine research institution, has simulated the specific path of nuclear pollution emissions, showing the important impact of nuclear pollution emissions on the environment. As shown in Figures 1, from the date of nuclear sewage dis-charge, the diffusion to most parts of the Pacific Ocean within 57 days will be affected by high radioactive substances. After 10 years, radioactive substances will spread throughout the entire ocean, and even reach all corners of the land with ocean currents and precipitation.

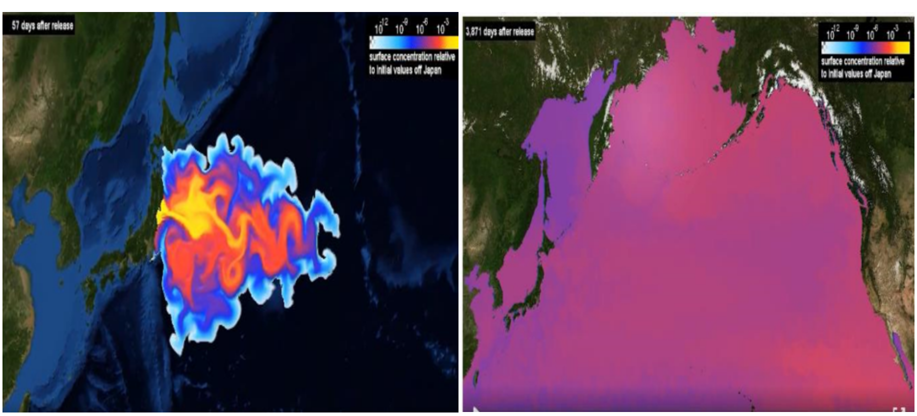


Figure 1. The diffusion range of nuclear pollution after 57 days and 10 years.

2.2. The Illegal Cost of Pollution Is Too Low

In 1985, the States parties to the ' London Convention ' agreed at the Ninth Consultative Conference on a moratorium on the marine disposal of low-level radioactive wastes resolution 21, and decided to maintain an indefinite moratorium on the dumping of low-level radioactive wastes and other radioactive materials at sea until the final research and evaluation. [4] In addition, the modalities for the disposal of nuclear wastes are clearly defined in a number of international regulations, such as ‘London Dumping Convention’, ‘Convention on the Safety of Spent Fuel Management’ and ‘Convention on the Safety of Radioactive Waste Management’, and ‘United Nations Convention on the Law of the Sea’, as shown in table 1. It is an internationally wrongful act for Japan to discharge nuclear waste into the sea in violation of international provisions and evade its obligations under international law.

Table 1. Modalities for the treatment of nuclear pollution and provisions of international law.

Types of nuclear waste	Principle of management	Processing mode	Countries	Provisions of international law
intermediate and low level radioactive wastes	Multiple Barrier principle	Near-surface shallow buried disposal or cave disposal	USA, France, Germany	‘Joint Convention' ‘London convention’ ‘treaty on the non-proliferation of nuclear weapons’
high level radioactive waste	Absolute security principle	deep geological repository	USA	
spent fuel[2]	Spent fuel treatment is carried out with reference to high-level radioactive waste			

[2] According to the provisions of relevant international law

However, due to the lack of operability and enforcement, the relevant international laws and regulations are only moral condemnation in dealing with Japan’s nuclear pollution, and there is no actual legal responsibility, which is not conducive to building a good international order. Low costs of illegality can also make such violations emulated by other countries, leading to increased marine pollution.

GEOMAR, a German marine research institution, has simulated the specific path of nuclear pollution emissions, showing the important impact of nuclear pollution emissions on the environment. As shown in Figures 1, from the date of nuclear sewage dis-charge, the diffusion to most parts of the Pacific Ocean within 57 days will be affected by high radioactive substances. After 10 years, radioactive substances will spread throughout the entire ocean, and even reach all corners of the land with ocean currents and precipitation.

3. Macroscopic Control Function of Government

The position announcement of environmental accounting and reporting adopted in 1999 makes the following definition of environmental cost: ‘Environmental cost refers to the cost of measures required to manage the impact of enterprise activities on the environment in accordance with the principle of environmental responsibility, and other costs paid by enterprises to implement environmental objectives and requirements.’ The definition clearly defines enterprise-centred and primary responsibility for eliminating environmental impacts.

Nevertheless, the environmental costs of different enterprises are different. For example, chemical enterprises, metal industries and other special industries have a high threat to environmental pollution, and they also consume high environmental control costs. [5] As shown in Table 2, the environmental costs of different industries are quite different, and the key pollutants to be treated are also different. In order to focus on solving the principal contradiction, it is necessary for the government to carry out overall planning and identify an industry in the area under its jurisdiction as the target to calculate and control the pollution of all enterprises in the industry and reduce environmental costs.

Table 2. Environmental costs of different industries.

Types of business	Hi-tech industry	Chemical enterprise	Pharmaceuticals industry	Metal industry
A	2574	22855.97	28809	24000
B	2344	21536.55	12483	30000 [3]

[3]Table 2 Based on CSMAR Database

In other words, from a macro perspective, government guidance is the premise and the government has the right and obligation to control and guide the environmental costs of society. Whether the enterprise' s environmental cost control method is appropriate or not directly affects the government' s cost expenditure. It includes the cost of pollution control, the cost of compensation for the general public, and the cost of formulating policies to restrict and punish enterprises.

From the perspective of cost and benefit in law and economics, the government save costs and reduce government spending is the general trend. So how to save costs, which requires the government to be based on the whole, draw lessons from successful experience and for enterprise environmental cost control policy, formulate corresponding policies in advance.

4. Macroscopic Control Function of Government

4.1. Improving Relevant Laws and Regulations on Environmental Cost Control

Our government must be committed to formulate environmental cost control regulations for enterprises, improving environmental cost accounting laws and regulations, and strengthening the administrative punishment of enterprises, which is the necessary premise to improve the current situation of environmental cost control in China's enterprises, because legal policies must play an important role in the actual behavior of enterprises.

Specifically, the coverage of China ' s current environmental regulations is narrow, the content and requirements of environmental cost control are not clear and environmental protection standards are not uniform. Therefore, it is necessary to improve from the above three directions. At the same time, it is necessary for the government to actively learn lessons from the laws and regulations of countries with similar successful experience and add the elements suitable for our country to the laws and regulations of our country in order to formulate specific environmental cost control criteria. By clarifying the significance of environmental cost control, and the disclosure of environmental cost control information, so as to better guide the practice of Chinese enterprises and achieve the win-win situation of enterprise economy and social benefits.

4.2. Strengthening Oversight Guidance

Perfect laws and regulations can not guarantee that enterprises can consciously abide by the relevant system. Therefore, relevant government departments need to increase guidance and supervision of enterprises. Implement real-time monitoring of enterprise environmental cost control, strengthen administrative penalties for violations, formulate more specific and detailed cost control requirements, timely and comprehensively guide enterprises to carry out environmental cost control. [6] In addition, the government can also reasonably use price control means and media supervision means to promote and establish the concept of green production of enterprises, and divide regions for centralized pollution control, so as to reduce the environmental cost expenditure of enterprises in the region.

4.3. Increasing Disclosure of Corporate Environmental Cost Information

The China Securities Regulatory Commission and the Ministry of Finance undertake the disclosure and supervision of relevant accounting information. Since the environ-mental cost of enterprises also involves environmental protection, and the statistical data and corresponding indicators of the National Environmental Protection Bureau are more convincing, the National Environmental Protection Bureau also shoulders the responsibility of information disclosure. In this way, it is not only conducive to spur enterprises to control environmental cost, but also conducive to the use of international platforms to improve corporate environmental awareness and encourage the government to comply with the trend to develop a sound control policy. Firstly, the information released by the enterprise should be reliable, rigorous, and timely. Secondly, professional staff should strengthen the research of information disclosure technology, and formulate relevant audit computer system. Furthermore, setting up a specific audit department in order to see whether the purpose of each environmental cost expenditure is reasonable and amount of funds is not enough.

4.4. Increasing Disclosure of Corporate Environmental Cost Information

Emission permit trading refers to the environmental protection department to formulate total emission control index, and according to the total control index, the environmental protection department issued emission permits to sewage enterprises. [7] According to the requirements of total amount control, pollutant emission enterprises must discharge ac-cording to the requirements of emission permits which can be traded between pollutant emission units. On the one hand, enterprises can't choose to control pollution emissions and offset their excessive emissions by purchasing the emissions of other enterprises. On the other hand, enterprises with residual emissions can sell their emissions to other enterprises and profit from it. Pollution permit trading is more flexible and effective than the traditional pollution charge system.

Emission permit trading is an effective means of market financing for environmental protection industry. The government can give indirect subsidies to enterprises by trading pollution permits to effectively control pollution, and give indirect penalties to enterprises with poor pollution control. Thus, encouraging enterprises to increase investment in environmental protection equipment and increasing demand for environmental protection products. In addition, the government can also raise funds for the environmental protection industry by auctioning permits.

5. Conclusions

Environmental cost control is a long-term and complex work. In order to better implement cost control, the government needs to make long-term efforts and give play to the role of macro guidance, learn from the lessons of Japanese government and enterprises' laissez-faire pollution, distinguish pollution control indicators in different industries, and improve measures in many aspects such as law, regulation, and system, so as to better promote the high-quality development of enterprises.

References

[1] Liu JG, Wang XY, Tan ZJ, Chen JH. A tripartite evolutionary game analysis of Japan's nuclear wastewater discharge. Ocean and Coastal Management. 2021; 214:61-70.
[2] Qiu J, Chen QM. Research on enterprise environmental cost control from the perspective of sustainable development. Frontiers in Economics and Management. 2021; 2:77-80.
[3] Gong YJ, Gu WH, Ren WF. Green energy economic efficiency and enterprise environmental cost control based on the internet of things. Security and Communication Networks. 2022; 2022:96-100.
[4] Xu XY. Analysis of enterprise environment cost control under haze background. Proceedings of 2016 5th International Conference on Social Science, Education and Humanities Research (SSEHR 2016) 2016, 331-334.
[5] Song CL, Han LR. The views and control of environmental cost. Journal of Modern Accounting and Auditing. 2010; 6(6):28-32.
[6] Florida R, Davison D. Gaining from green management: environmental management systems inside and outside the factory. California Management Review. 2011; 43,64-84.
[7] Amelia DK, Cyril P. The joint convention on the safety of spent fuel management and on the safety of radioactive waste management. The International and Comparative Law Quarterly. 1998; 47(2):411.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230207

Research on Innovative Strategies for Online Education Management in Universities in the Era of Big Data

Wen HUANG[a,1] and Yonghui GUO[b]
[a] *School of Literature and Journalism, Xiangnan University, No. 889, Chenzhou Avenue, Chenzhou 423000, China*
[b] *School of Physics and Electronic-Electrical Engineering, Xiangnan University, No. 889, Chenzhou Avenue, Chenzhou 423000, China*

Abstract: In the era of the Internet, the continuous development of information technology represented by 5G, big data, artificial intelligence, etc. has driven the transformation of university education from traditional offline education mode to digital, networked, and intelligent online education mode. Online education in universities has become an important way to promote educational reform and innovation. However, in the process of Digital transformation of education, there are still many problems such as solidification of teaching management mode, insufficient integration of online resources, lack of intelligent education practice platform, etc., so it is urgent to improve the online education management ability of colleges and universities. In the future, university education management needs to use big data tools to improve management level, explore educational resources, unify teaching and learning, and achieve the transformation and upgrading of online education management mode.

Keywords: Big data, universities, online education, administration

1. Introduction

Big data relies on the Internet as the basic data material, with the characteristics of scale and diversification, and can provide the required large amount of data in a short period of time. Education is constantly updating and improving with the development of the times. The arrival of the big data era has brought about significant changes in modern education, and big data will inevitably become the driving force for education reform. Promoting and improving the current online teaching management mode in universities has become an important issue that needs to be urgently solved [1-2]. At present, there is a shortage of professional talents in the field of big data technology in China, and most university teaching management personnel have not fully mastered the relevant technology, resulting in a relatively poor application of big data [3-4]. How to

[1] Wen Huang, School of Literature and Journalism, Xiangnan University, No. 889, Chenzhou Avenue, Chenzhou 423000, China; E-mail: 450558724@qq.com.
This research was funded by: The Mixed Teaching Reform and Practice of Huxiang Intangible Cultural Heritage Blended with Ancient Chinese Literature under STEAM Education Concept (Grant Number: HNJG-2022-1083); Research on Optimizing the Teaching Content of Motion Control System Course under the OBE Concept (No. 22).

effectively apply big data in teaching management, help educators establish big data thinking and master relevant technologies, analyze scientific, reasonable, and efficient teaching management methods for college students' data information, and improve the quality of teaching management has become an urgent issue to be solved.

2. Advantages of Online Education Management in Universities in the Era of Big Data

2.1. Improving Teaching Efficiency

Applying big data and artificial intelligence data to teaching management can break traditional inherent thinking and open up new prospects. Compared to previous work modes, this emerging method of combining artificial intelligence and big data has led to a rapid improvement in work efficiency. The application of big data technology can fully enhance the teaching and management capabilities of schools, improve the quality of talent cultivation, and become an irreversible development trend in today's era [5]. From the current development status, big data is gradually being applied to different teaching types, as shown in Table 1.

Table 1. Data source types of big data education.

Learning categories	APP	Data source
Self study	Wisdom Study Companion, Discussion and Answer	Homework data, Exam data
Homework Q&A	Homework Help, Afan Question	Homework data, Question bank data
Language learning	Fluent in English speaking, proficient in a hundred words	English speech database, Word database
Teaching	Rain Classroom, Mu Ke	Interactive data between teachers and students

Teaching managers use different teaching platforms to guide teachers and students to apply online resources to self-study, homework, and classroom teaching. And through data sources such as homework, exams, knowledge, etc., grasp the teaching level of teachers and the learning situation of students, in order to further improve teaching management.

2.2. To Meet the Teaching Requirements of Individualized Teaching

By utilizing technologies such as big data processing and deep reinforcement learning, learners' digital portraits can be calculated and depicted. Through dynamic data learning, students' psychological states and behavioral patterns can be calculated, providing personalized teaching services for each student. Teaching managers need to adapt faster to the ever-changing teaching environment, fully understand students' learning behavior through the analysis and screening of teaching data information, so as to achieve individualized teaching and propose targeted teaching plans for different students, in order to improve the quality of talent cultivation and promote students' comprehensive development.

2.3. Reforming the Teaching Evaluation Model of Curriculum

The application of big data in teaching management can bring great opportunities for the reform of teaching quality evaluation, allowing students to fully absorb the knowledge and energy of different disciplines and apply them to the deep learning process, promoting students' comprehensive development. In terms of evaluation standards, evaluation methods, and content, the teaching evaluation model in universities can innovate by using big data to collect a large number of students' learning situations, and analyzing data through computer and other technologies, ultimately forming visual teaching suggestions [6-7]. Based on the above concept, a procedural evaluation system is formed, as shown in Figure 1.

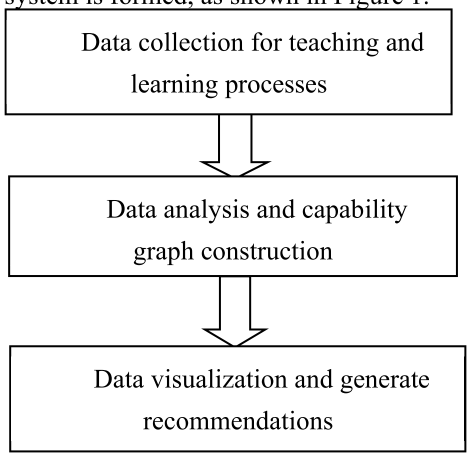


Figure 1. Process evaluation system diagram of big data application.

3. The Problems in Online Education Management in Universities in the Era of Big Data

3.1. Insufficient inTegration of Teaching Management and Technology

Chinese universities have integrated big data technology into their teaching management, but generally remain focused on data integration, such as counting the number of students on campus, registering for exams, and counting the number of students in each score segment. The application of technology is relatively limited. After some universities introduced big data technology into teaching management, the application direction did not align with their actual needs, resulting in a small scope of technology application and poor effectiveness. In the teaching management work of universities, it is necessary to break away from the teaching management itself and simply rely on data for teaching management and innovation. Among numerous data sources, blindly educating college students through data is not conducive to achieving teaching effectiveness. To make good use of data information, as a teaching management worker, it is necessary to have strong data screening ability, high level of informatization, perfect integration of teaching management and emerging technologies, and achieve innovation in management methods and means. In addition, there is insufficient emphasis on information security protection, a lack of awareness of security management, and insufficient protection of data information, which affects the stable operation of the system and disrupts the order of teaching management.

3.2. Lack of Service Awareness in Traditional Teaching Management Models

Some university teaching management systems focus on the management of basic information for teachers and students, neglecting the service function for teachers and students. Only some of the daily teaching affairs of teachers and students can be completed in the system, and more still need to be solved offline. For students, they can query their grades, select courses, and browse their schedules in the system, but they cannot apply for leave, leave cancellation, or modify their academic status. The traditional management model, which sets up management systems from the perspective of managers, neglects the user experience of teachers and students, hinders the maximization of teaching management functions, and also makes big data technology unable to be applied.

3.3. Insufficient Information Technology Vapabilities of Teaching Management Workers

Decision makers in the management process of universities are not aware of the importance of big data technology for educational curriculum reform, making it difficult for various departments in universities to have a correct understanding of the importance of big data. This skeptical attitude at the decision-making level directly raises the threshold for big data to enter the education industry, blocking its development process. This reflects a prominent issue in the application of big data in university teaching management, and is also the main reason for the narrow application field of big data in university teaching management.

4. Strategies Adopted for Innovation in Online Education Management in Universities in the Era of Big Data

4.1. Deepen the Integration of Teaching Management and Big Data Technology

In the context of big data, using big data technology to construct relationship models, identifying relevant relationships through data, and predicting college students' thoughts and behaviors is beneficial for mastering student dynamics in the first time and better teaching management. Firstly, change the existing teaching management methods, implement teaching management prediction models, and establish predictive thinking for teaching managers. Education managers should establish data awareness, scientifically establish student behavior models using mastered data, and analyze students' thinking patterns and behavior predictions based on scattered data and information. By analyzing the results, a reasonable talent development plan, course arrangement, and credit point setting can be formulated to better understand how to manage teaching. Predict students' future behavior based on data, understand their future development trends, and respond quickly and skillfully to emergencies. Secondly, strengthen the ability to collect and analyze student data information. Data collection and analysis are the basic tasks of educational informatization, and the efficiency and accuracy of data collection will directly affect the effectiveness of data analysis and modeling. Furthermore, it is necessary to strengthen risk management and information assessment. By collecting and processing data, not only can educational resources be excavated from the data, but also early warning judgments can be made through data analysis. Finally, strengthen technical support to ensure data and information security.

4.2. Innovative Teaching Management Mode

Teaching management is a process of coordinating and managing various elements of the teaching process, making it operate effectively and improving efficiency. At present, from the current application status of big data in teaching management, it cannot be separated from infrastructure construction. Infrastructure construction is a very important organic part of the expansion of big data in the field of education. For service innovation, schools should inject the educational concept of comprehensive quality and comprehensive development into students' education work, and promote infrastructure construction, in order to better establish service functions and teaching quality evaluation paradigms. The basic education work of students cannot be separated from the guarantee of infrastructure, in terms of investment and construction of infrastructure guarantee [8-9]. According to the research results of the Zhongyan Puhua Industry Research Institute, it is predicted that the scale of China's education industry informatization market will gradually increase, as shown in Figure 2.

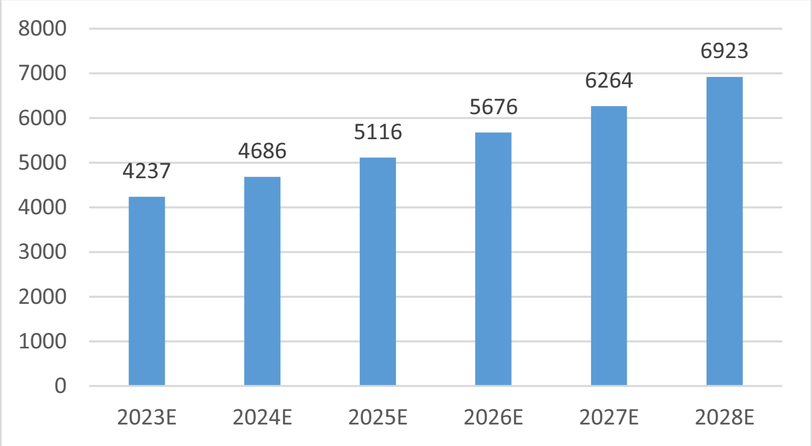


Figure 2. Supply Forecast for China's Education Industry from 2023 to 2028 (100 million yuan).

In the era of the Internet, the construction of resource sharing in university teaching management under the background of big data is an inevitable trend. This helps to achieve a teaching management model of information integration in university teaching management systems and integration of university information equipment, providing means for exploring teaching laws and opportunities for the construction of management decision-making levels in universities [10]. According to the data released by big data trading, the application of big data in the education field does not have an advantage, as shown in Figure 3.

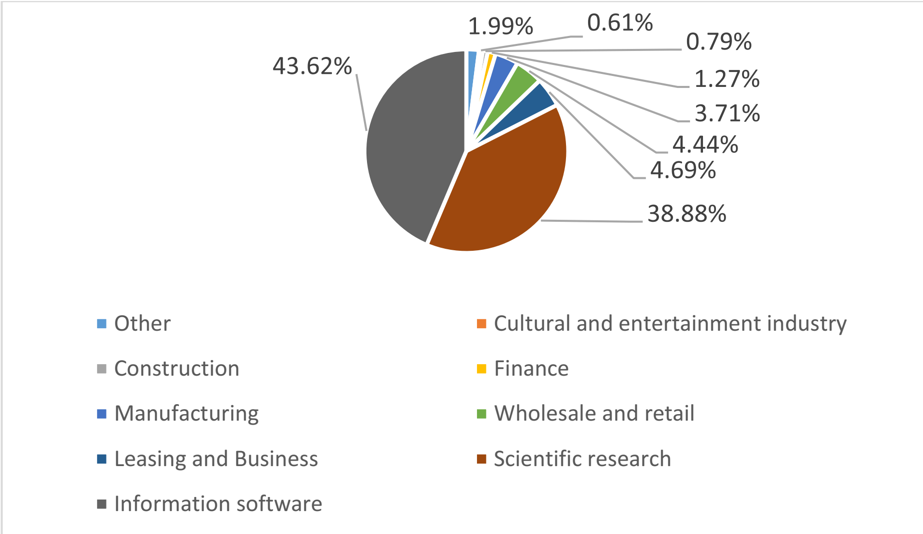


Figure 3. Proportion of Big Data Enterprises by Industry.

Universities should establish correct values in a timely manner, increase the investment budget for introducing big data into infrastructure construction, and improve the service quality and level of school teaching. This is an important development requirement for university teaching management in the new era.

4.3. Strengthening Teaching Management and Cultivating Teachers

Innovative and efficient teaching management models can orderly carry out teacher training activities in the process of education and teaching reform and comprehensive management work. This can enable teachers to fully utilize big data technology to achieve curriculum education and comprehensive quality education for students, which will greatly benefit teaching quality. The importance of big data for education lies in promoting the modernization of teachers and the informatization of teaching. Through the application of information technology concepts, tools, environments, and methods, promote teachers to adapt to changes in the times and achieve modernization and transformation of teachers [11]. The information technology elements of Chinese university teachers are shown in Figure 4.

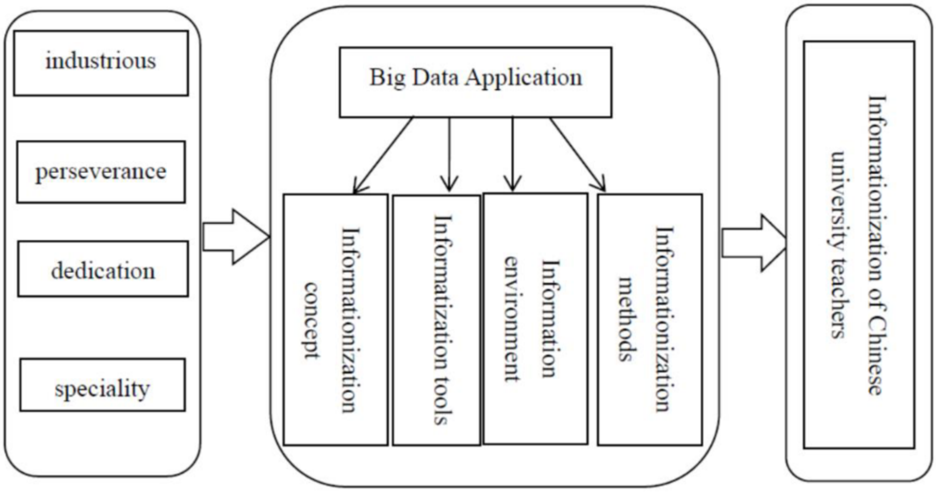


Figure 4. Elements of Informationization for Chinese University Teachers.

5. Conclusions

Traditional classroom education in Chinese universities should adapt to the changes of the times and make reasonable use of big data to strengthen teaching management. The innovative development of online education must be combined with the advantages of the times, and the integration of big data technology and artificial intelligence technology has become an inevitable trend in the reform of education methods and education systems. Starting from the teaching quality management mechanism, big data organization framework, and smart teaching platform, we aim to innovate the online education model, in order to achieve healthy, stable, and sustainable development of the online education model, and create a free and flexible intelligent teaching method for citizen education.

References

[1] Sun JP, Wang XT, Zhen XY. On the innovation of online education mode in the era of "internet plus". Journal of Shanxi radio & TV University. 2021; 125: 36-40.
[2] Yang XM, Yang KF, Zhang YR. The potential dilemma and response strategies for the effective implementation of the "double reduction" policy: Also on the construction of the public online education service system. Journal of Sichuan Normal University (Social Sciences Edition). 2021, 48(6): 53-61.
[3] Shao LF. Research on online teaching management in vocational colleges under the background of big data. Technology Style. 2022, 14(5):25-27.
[4] Zhang J. Big data: A new way for vocational colleges to modernize teaching management. Journal of Educational Modernization. 2019, (101):222-223.
[5] Fang MS, Liu BB. Innovation in university teaching management under the background of big data. Industry and Technology Forum. 2021, 20(24): 283-284.
[6] Gong YQ. Research on innovation of teaching management in universities under the background of big data. Research on Higher Education of Nationalities. 2022, 10(4):88-92.
[7] Yang SN, Li GR. Research on online teaching management in vocational colleges under the background of big data. Wireless Internet Technology. 2022, (3):134-135.

[8] Cui Y. Analysis of innovation in teaching management in universities under the background of big data. China Management Informationization. 2020, 23(17):222-223.
[9] Chen L, Ren PP, Zhang WM. A new perspective on educational innovation and development in the post pandemic era and China's outstanding exploration: Reflections on attending the 2020 global conference on artificial intelligence and education big data. China Educational Technology. 2021, (412): 1-9.
[10] Mao YB, Li XY, Zhang L. Research on the improvement of online education quality in the digital transformation of education. Digitalization of Education. 2022, (428):38-42.
[11] Song M. The ongoing and future tense of online education. Media Culture. 2022, (7):62-65.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230208

Difficulties and Strategies of Digital Transformation of Lifelong Education

Wen HUANG[a,1] and Yonghui GUO[b]
[a] *School of Literature and Journalism, Xiangnan University, No. 889, Chenzhou Avenue, Chenzhou 423000, China*
[b] *School of Physics and Electronic-Electrical Engineering, Xiangnan University, No. 889, Chenzhou Avenue, Chenzhou 423000, China*

Abstract: With the development of China's economy and social progress, China's education industry has ushered in huge development opportunities. The development of high-tech has accelerated the integration and utilization of educational resources, providing opportunities for the development of lifelong education. Promoting the transformation and upgrading of traditional lifelong education to Digital transformation is an important guarantee and an inevitable trend for China to pursue the sustainable development of lifelong education. By analyzing the changes in the current educational environment, this paper explores in detail the challenges faced by the Digital transformation and upgrading of lifelong education and the development strategies, which will play a certain reference role in the upgrading and development of Digital transformation of lifelong education.

Keywords: Lifelong education, digitization, transformation, informatization

1. Introduction

China is currently in the process of continuously developing its economic development level, and lifelong education is gradually entering the public eye from the periphery. The popularization and application of a series of information technologies such as big data, the Internet of Things, virtualization, and cloud computing in the field of education in China have prompted continuous updates and development of educational concepts, environments, and methods [1-2]. The Digital transformation of lifelong education is jointly determined by economic development, social changes, technological innovation and other factors, and ultimately aims to improve the quality of the people, improve their cultural level, and achieve the goal of free and personalized development of education [3-4]. In the new era, to implement the digital strategic action of lifelong education and accelerate the transformation and intelligent upgrading of lifelong education, it is necessary to explore the current problems and solutions of the Digital transformation of lifelong education.

---

[1] Wen Huang, School of Literature and Journalism, Xiangnan University, No. 889, Chenzhou Avenue, Chenzhou 423000, China; E-mail: 450558724@qq.com.
This research was funded by: The Mixed Teaching Reform and Practice of Huxiang Intangible Cultural Heritage Blended with Ancient Chinese Literature under STEAM Education Concept (Grant Number: HNJG-2022-1083); Research on Optimizing the Teaching Content of Motion Control System Course under the OBE Concept (No. 22).

2. Current Situation of Digital Transformation of Lifelong Education in China

2.1. The State Vigorously Promotes the Digital Transformation of Lifelong education

In recent years, China's policy documents on Digital transformation of education have been issued, which provides a better policy environment for promoting China's education transformation and upgrading to a new stage. In December 2021, the "Fourteenth Five Year Plan" National Informatization Plan proposed to "carry out lifelong digital education", and pointed out the direction of Digital transformation of lifelong education from infrastructure, digital resources, teaching reform and other aspects [5]. By May 2023, 29 provinces and cities in China have issued relevant policies on promoting the Digital transformation of education and accelerating the process of education informatization.

2.2. Existing Methods of Digital Transformation of Lifelong Education

The integration of new technology and education has promoted the continuous strengthening of education informatization in China, and the public's understanding of Digital transformation of education is deepening. Innovative lifelong education initiatives are constantly emerging and leading educational reforms, such as the O2O learning method, which enables learners to achieve synchronous online and offline teaching; U-Learning, utilizing information technology to provide intelligent learning tools that are not limited by time and location; MOOC and other online learning platforms offer high-quality courses for free and provide a large amount of learning materials. In addition, the implementation of various teaching resource sharing methods such as micro classes and flipped classes provides convenient conditions for realizing the Digital transformation of lifelong education [6]. According to the online questionnaire survey results, it can be seen that Chinese citizens tend to use online learning platforms, as shown in Figure 1.

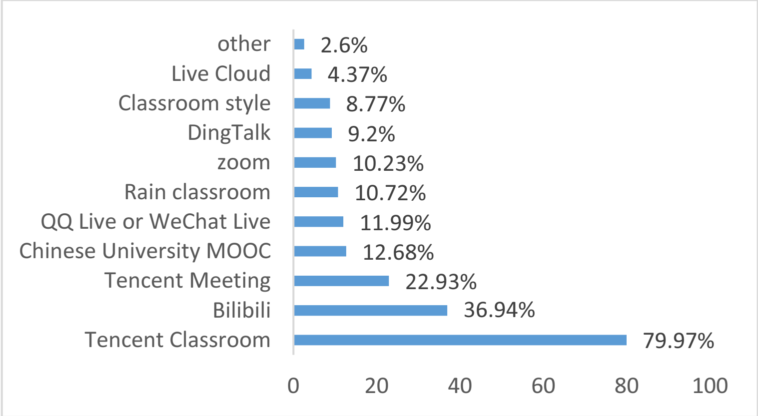


Figure 1. Chinese citizens tend to use online learning platforms.

According to the survey, the platforms that Chinese citizens most tend to use are Tencent Classroom, Tencent Conference, Love Classroom, QQ Live, and Rain Classroom.

2.3. Achievements of Digital Transformation of Lifelong Education

China attaches great importance to the construction of lifelong education learning resources, especially the digital and networked transformation and upgrading of learning resources. New generation technologies, such as artificial intelligence, big data and 5G, enable learners to acquire new capabilities and are constantly integrating with education [7]. Technology enabled education will make breakthroughs in lifelong education, update the concept of lifelong education with new generation information technology, transform the education model, and promote new forms of innovative development, such as continuing education for academic qualifications, non academic education training, elderly education and community education,as shown in Figure 2.

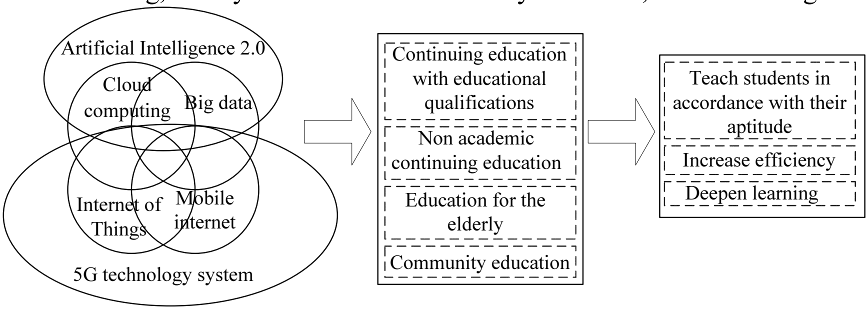


Figure 2. The Relationship Between Next-generation Information Technology and Lifelong Education.

The lifelong education network platform can provide more diverse and high-quality courses to everyone, which is the pursuit of educational fairness.

3. Difficulties in Digital Transformation of Lifelong Education

The Digital transformation of lifelong education is not easy. This paper discusses the typical difficulties in the Digital transformation of lifelong education from three aspects: infrastructure construction, education subject recognition, and the quality of talent and technology integration, as shown in Figure 3.

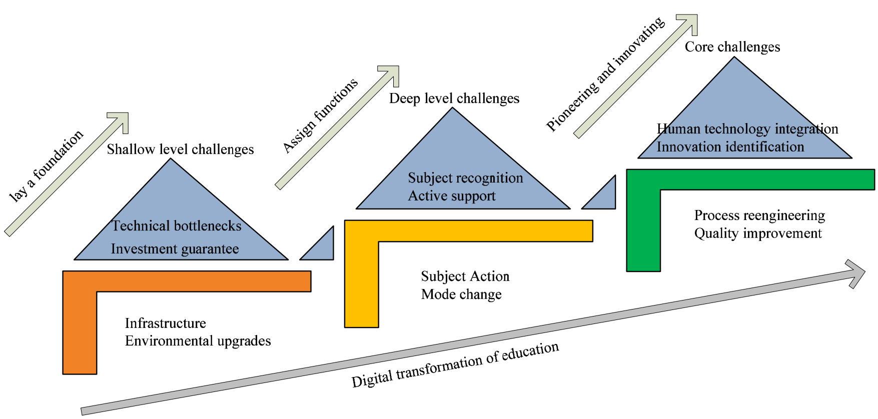


Figure 3. Typical problems in the Digital transformation of lifelong education.

3.1. Insufficient Infrastructure Construction

Infrastructure construction is the core force of educational transformation in the digital era, and an important strategic measure to accelerate educational modernization and build an educational powerhouse. There are still some deficiencies in the infrastructure construction of China's lifelong education and Digital transformation. The most critical element is the construction of digital education resource platform. At present, the digital high-quality education resources that can effectively meet the needs of the public are still in a state of long update cycles, insufficient planning and design, and the mechanism for sustainable quality improvement still needs to be improved [8-9]. The high-quality digital education resources and digital education platforms that meet the needs of lifelong education urgently need to be effectively strengthened.

3.2. Insufficient Identification of Educational Subjects

The Digital transformation of lifelong education also has the phenomenon of not deep understanding of the concept and insufficient theoretical research. Lifelong education digital educators do not yet have the awareness to fully utilize digital technology and big data to change traditional teaching content and methods, and achieve digital thinking leading the transformation of education and teaching. For example, the understanding of the connotation of Digital transformation of education and the concept of leading the modernization of education is not deep, the top-level design is not enough, the systematic methodology is still lacking, and the rich typical experience of successful transformation is also lacking, and the benefits of transformation are not very obvious.

3.3. The Quality of Talent and Technology Integration needs to be Improved

In the new era, China's goal of accelerating the construction of a high-quality education system and a powerful country with human resources forces the Digital transformation of education. The purpose of lifelong education is to realize the equalization of education public services, accelerate the construction of high-quality education system, and urgently need education Digital transformation enabling education. The new generation of digital technology and the Digital transformation of the economy urgently need education to cultivate innovative talents needed by the society and improve the quality of human resources to meet the needs of future international competition and building a powerful country in human resources [10-11].

4. Strategies for Digital Transformation of Lifelong Education

Based on the innovation Ecological systems theory and the technology organization environment framework, an analysis model for the driving factors of educational Digital transformation is proposed, as shown in Figure 4.

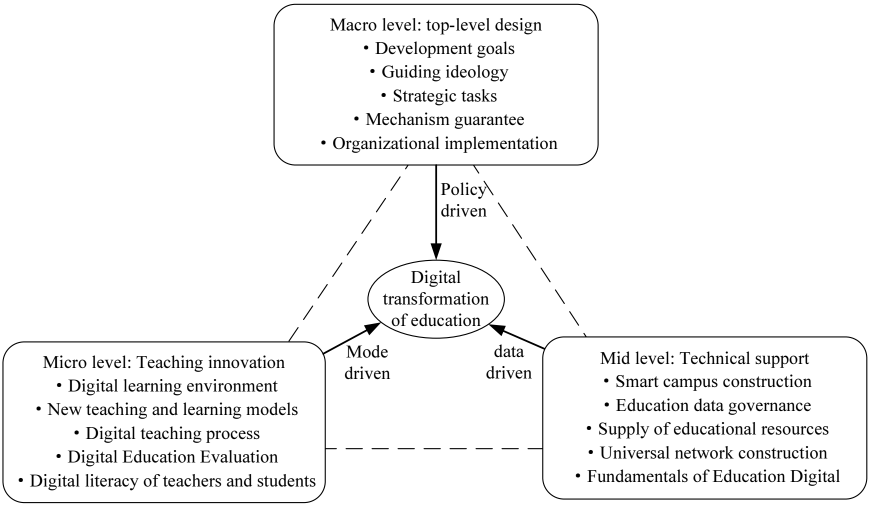


Figure 4. Analysis model of driving factors for Digital transformation of education.

4.1. Promoting the Infrastructure Construction of Lifelong Education Platforms

Deeply apply 5G, artificial intelligence, big data, virtual reality and other new generation digital technologies, accelerate the promotion of new digital infrastructure in information networks, platform systems, digital resources, smart campuses, innovative applications, trusted security and other aspects, and innovate the support system for Digital transformation of education. For example, the use of AI technology to promote the Digital transformation of teaching, the use of AI knowledge mapping technology support to achieve the Digital transformation of course materials, and the use of AI learner profiling technology support to achieve the Digital transformation of student learning difficulty classification research. Rural and urban areas jointly build "cloud classrooms", carry out remote online teaching, connect urban and rural areas with one screen, integrate and interact online and offline teaching, and share high-quality digital teaching resources. As shown in Table 1.

Table 1. Functions and Characteristics of Some Online Education Platforms in China.

Platform Name	Operator	Uptime	Function and characteristic
CCtalk	Hujiang Education	2012	Based on the learning center, it has achieved various functions such as online question answering, courseware display, voice interaction, whiteboard writing, and screen sharing, covering Android/iOS, iPad, PC, and Mac platforms. In 2019, V7.5.6 version was launched to further optimize the learning experience.
DingTalk	Alibaba Group	2014	By establishing nail groups, various online classroom functions can be achieved, using three live streaming modes: camera mode, screen sharing mode, and professional mode. The professional mode of group live streaming can be set to add images, set cameras, share desktops, play videos, insert images, capture windows, and so on. The platform provides PC, Web, and mobile versions, supporting file transfer between mobile phones and computers.

Tencent Classroom	Tencent Education	2014	A large number of high-quality educational institutions and famous teachers are gathered here. There are many online learning boutique courses, such as vocational training, Civil service examination, TOEFL IELTS, certificate examination, oral English, primary and secondary education, to create a classroom for teachers to teach online and students to interact in a timely manner.

The most commonly used learning software in China can provide a good platform for the Digital transformation of lifelong education. In the future, it will continue to integrate technology to build a dedicated learning network platform for lifelong education.

4.2. Improve the Digital Collaborative Governance System

Deepen the application of digital technology and promote systematic changes in education. Smart education is based on the application of intelligent technologies such as artificial intelligence, big data, and the internet in the field of education, and builds an intelligent learning atmosphere through methods and means such as platforms, resources, and data. This educational model provides personalized demand solutions for lifelong education, and guarantees the cultivation of talents with innovative practical abilities, interpersonal communication and cooperation abilities, high-level thinking abilities, digital literacy and skills in society. In short, it is to use the wisdom of education to cultivate intelligent people.

4.3. Strengthen the Construction of Digital Talent Team

The digitalization level of the teaching staff affects the implementation of the digital education strategy. To cultivate professional digital talents and enhance the digital literacy of teachers and students, it is required that teachers and students can use certain information technology means and methods in the digital environment to quickly and effectively discover information, integrate information, extract information, and communicate and store it effectively. Improving the digital literacy and digital application ability of teachers and students is the key to realize the Digital transformation of education. Digital talents usually need to possess three elements: one is to master digital technologies related to computers, big data, artificial intelligence, communication, etc., which is the basic requirement for digital talents; The second is to be able to apply digital technology in a certain field or scenario, which is a prerequisite for digital talents to play a role; The third is to be able to keep up with technological and industrial development trends, which is a professional requirement for digital talents. The promotion of lifelong education Digital transformation needs talents. Every step of digital development can not be separated from talents. The development of digitalization in every industry, industry and enterprise is the result of talents' efforts. Digital transformation is not only a technological transformation, but also a talent transformation fundamentally.

5. Conclusions

The Digital transformation of lifelong education needs to continuously strengthen the deep integration between science and technology and education, accurately locate the construction goals and tasks, develop a complete workflow of resources from construction to integration, development and utilization, and standardize the technical standards and development platform for the construction of digital learning resources. By constructing the digital concept of lifelong education, we aim to enhance the modern resource sharing and utilization of digital, networked, and integrated learning resources, create a scientific digital teaching resource management approach, and promote the development of lifelong education.

References

[1] Xiong J, Zeng H, Zhang YD. Digital transformation and intelligent upgrading: research on the construction of "Internet plus" new universities. Chinese Journal of ICT in Education. 2023, 29(2): 64-78.
[2] Yang LJ. Research on the construction of open university education system in comprehensive reform. Journal of Shaanxi Open University. 2022, 24(2):5-13.
[3] Zhu ZT, Hu J. Education digital transformation: future oriented education "transgenic" project. Open Education Research. 2022, 28(5): 12-19.
[4] Yan P. Strategic positioning and development strategies of open universities under the new situation. Journal of Ningbo Open University, 2021,19(4): 5-10.
[5] Zeng H, Hong YA. Research on the practice of Guangzhou intelligent lifelong learning public service model. Journal of Guangzhou Radio and Television University, 2019,19(5): 18-22108.
[6] Zeng H, Xiong Z, tang JP. Digitization, standardization, specialization, and branding: model selection for the development of Guangzhou digital learning port. China Distance Education, 2013(4): 11-15.
[7] Zhu ZT, Zheng H, xie LJ. Demand analysis and action suggestions for the digital transformation of education empowered by new infrastructure. Open Education Research, 2022, 28(2): 22-31.
[8] Zheng XD, Zhou ZH. Three questions about new education infrastructure: What is the foundation? Where is the new one? How to build it? Research on Electronic Education, 2021, 42(11): 42-47.
[9] Chen H, Xiong Z. Research on the digital strategy of higher education. China Higher Education. 2022(9): 7-9.
[10] Zeng H, Xiong Z. Exploration of the operation mode of "similar e-commerce" in digital continuing education. Journal of Guangzhou Radio and Television University. 2012,12(6): 14-17.
[11] Zeng H, Jia XX, Chen PD. Research on the open service model of digital learning resources for continuing education in open universitie. China Electronic Education. 2014(11): 50-54.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230209

Effects of Multi-Factors on Biomass Gasification Based on Big Data

Zehua WANG[a1], Lihua ZOU[a], Dongrui CHEN[a] and Sunzhi YONG[b]
[a] *Library, Shaoyang University, Hunan Province, 422000, China*
[b] *School of Energy Engineering, Yulin University, Shannxi Province, 719000, China*

Abstract. Bio gas is a kind of clean energy, which is often produced by biomass gasification. At the present work, the effects of multi-factors on the volume fraction of gas were analyzed numerically based on big data, and these factors included temperature, pressure, biomass species and various gasify agents. Moreover, the volume fractions and mass fluxes of gas was discussed. Pressure and temperature were set at the ranges from 1MPa to 6MPa and from 400°C to 1000°C, and the gasify agents selected H_2O, H_2O-O_2, O_2, CO_2 and H_2. The results show that the highest yields of gas were the pine sawdust with the highest content of C, H and H_2O since the yield of gas depend on the component contents of biomass. Pressure has a reverse effect on the variations of gas volume fractions compared to temperature. As pressure rose, the volume fractions of both H_2 and CO reduced, while the volume fractions of both CH_4 and CO_2 ascended. On the contrary, the volume fractions of both H_2 and CO increased, while the volume fractions of both CH_4 and CO_2 decreased with the increasing of temperature.

Keyword: Biomass gasification, gasify agent, temperature, pressure

1. Introduction

The emission generated by the combustion of fossil fuels have brought serious environmental pollution. As the exhaustion of fossil fuel and the calls on environmental protection, it is popularized for bio-gas as a kind of clean fuel due to its higher thermal value and lower emission. Valente et al. [1] checked the suitability of various gas production options based on the life-cycle. Gas production from biomass gasification performed significantly better than that from the steam methane reforming in environmental protection. Many scholars have studied the effects of operational parameters and gasify agents on biomass gasification at home and aboard. Fremaux et al. [2] run an experiment of the sawdust gasification with steam in a fluidized-bed. At 700°C and a residence time of 40min, as the addition of steam increased, the volume

1 Zehua Wang, Library, Shaoyang University, Hunan Province, 422000, China; E-mail: 2387994363@qq.com.
This work was supported by the key scientific research project of Hunan Provincial education department [Grant numbers 20A448], Hunan Province Natural Science Foundation of China [Grant numbers 2021JJ30633] and [Grant numbers 2022JJ50246], the outstanding youth scientific research project of Hunan Provincial education department [Grant numbers 22B0747], the guiding project for science and technology innovation of Shaoyang [Grant numbers 2022GX4047], the Project of National Natural Science Foundation of China [Grant numbers 52064048].

fraction of gas ascended, including H_2, CH_4 and CO_2. They reported that the increment in temperature surged slightly H_2 production, while reduced the tar content. Wang et al. [3] carried out an experiment of the sawdust gasification with steam in a fixed-bed, as residence time ascended from 8 min to 34 min and the temperature increased from 550 °C to 850 °C, the mass fraction of residual carbon reduced from 32% to 20%, the mass fraction of gas rocketed.

However, the present researches adopted the limited data to carry out their studies, the obtained results were only suit for given conditions. As the big data was applied widely, it is necessary to improve the application ranges of our findings under the help of the big data. Therefore, at the present work, on the basis of the big data, the numerical simulation was proceeded to analyze the effect of multi-factors on biomass gasification. The calculation work was done by Aspen Plus software, the conditions was selected based on the principle of big data.

2. Numerical Simulation

2.1. Principle of Biomass Gasification

Fig.1 shows the processing and principle of biomass gasification. Biomass first experiences drying and pyrolysis, forming gas products (CO, H_2, CH_4, CO_2, H_2O), liquid products (tar, liquid acid, macromolecule compound) and solid products (carbon and ash). Next, the pyrolysis products continue to be splitted, reformed, oxidated and reduced. Usually, the required heat during drying and pyrolysis will be satisfied by combustion reaction in biomass gasification, and main reactions was shown in Table 1.

Table 1. The main reactions in the Gasifier

Chemical reaction	ΔH (kJ.mol^{-1})	Name
Biomass→H_2O+H_2+CO+CO_2+tar+hydrocarbon+coke	—	pyrolysis
C_mH_n+H_2O→CO+H_2	Endothermic reaction	Reforming between CmHn and steam
C_mH_n+CO_2→CO+H_2	Endothermic reaction	Reforming between H_2 carbon and CO_2
CO+H_2O=CO_2+H_2	-41	Water and gas reaction
3C+4H_2O=2CO+CO_2+4H_2	+353	coke and steam Gasification o
C+O_2→CO_2，C+CO_2=2CO	-394, +172	Coke gasification
CO+3H_2=CH_4+H_2O CO_2+4H_2=CH_4+2H_2O	-207, -166	Methanation reaction

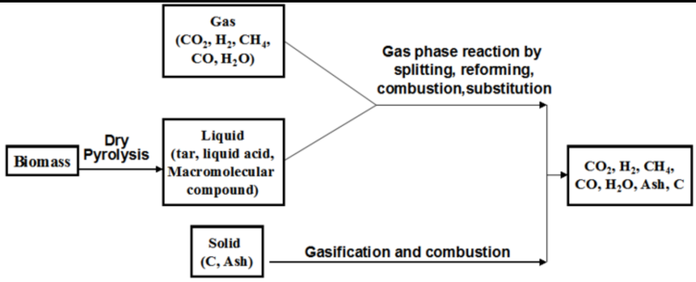


Fig. 1 Schematic diagram of the processing and principle of biomass gasification.

2.2 Calculated Method

Aspen Plus software has a successful application in coal chemical industry as a reliable method to simulate biomass gasification due to the similar components between biomass and coal. The assumption was done before the starting calculation[4].

(1) Gasifier is in a stable state without pressure and temperature gradients;

(2) O, H, S and N are transformed into gas, while C has a partial transformation;

(3) Ash is viewed as a kind of inertia substance without gasification;

(4) An instantaneous complete mixture occurs between gas and biomass;

(5) All gas phase reactions occur and reach an equilibrium state quickly;

A calculating processing was shown in Fig.2. Three modules were used, including RSTOIC (RS), RGIBBS (RG) and SSPLIT. Biomass was decomposed into the molecules in RS where realized a partial oxidation and gasification reactions, and then the produced heat was sent to RG for providing the oxidation reaction with uneven temperature. At the same time, various gasify agents were sent into RG, and both gas products and ash were separated with SSPLIT, and then entered to SEP for removing H_2O. Through the series of processing, the dried syngas was obtained.

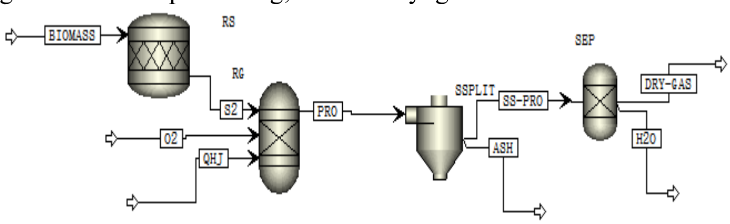


Fig. 2. Calculating processing of the biomass gasification model in Aspen Plus.

3. Comparison of Calculated Results to Experimental Data

In order to verify the reliability of Aspen plus simulation method, the experimental operational parameters in the reference [5] were simulated numerically. According to the experiment from the reference [6], the straw of 3g was used, and its components were shown in Table.1. The steam mass flux of 1.033g/min and the temperatures of 750℃, 800℃, 850℃, 900℃, 950℃ and 1000℃ were given. Fig. 3 shows a comparison of the calculated results to experimental data of H_2, CO, CO_2 and CH_4.

Table 2. Industry analysis and element analysis of sawdust.

Industry analysis (%)			Element analysis (%)					Q_{net}
FC_d	V_d	A_d	C_d	H_d	N_d	S_d	O_d	(MJ/kg)
15.33	84.1	0.54	46.73	6.54	1.71	0.95	43.47	18.85

It can be seen from Fig.3, as the temperature increased from 750℃ to 1000℃, the errors between the calculated results and experimental data ranged from -8% to 2% for the volume fraction of H_2, from -24% to 42% for the volume fraction of CO, from -52% to 27% for the volume fraction of CO_2 and from -28% to 95% for the volume fraction of CH_4.Their error ranged from -40% to 19% at 750℃, from -4% to 17% at 800℃, from -50% to 52% at 850℃, from -41% to 77% at 900℃, from -43%to 88% at 950℃ and from -42% to 95% at 1000℃. So, errors were neglected at from 700℃ to 800℃. It is mainly because the content of CH_4 reduced sharply and up to zero over

800°C, leading to an intensive increment in the error of CH_4. However, the calculated results at all temperatures are in agreement with the experimental data for the volume fraction of H_2.

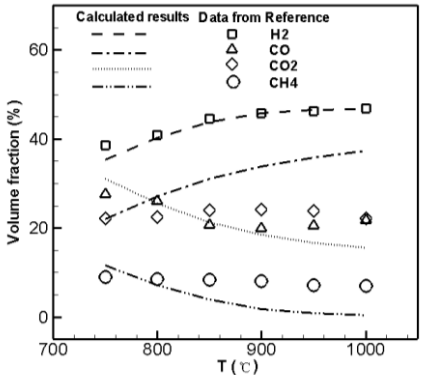


Fig. 3. Comparison of experimental data to calculated results.

4. Gasification Processing of Various Biomass

Biomass can be divided into herbaceous and woody plants with the macro-molecule polymer that consisted of C, H and O elements. Here, the herbaceous biomass selected the rice husk and corn straw, and the woody biomass considered the pine sawdust. Table.3 shows their industry analysis and element analysis. Aspen plus software was applied to simulate their gasification processing.

Table 3. Industry analysis and element analysis for three kinds of biomasses from reference [4].

Base	Component	Rice husk	Rice straw	Pine sawdust
Industry analysis (%)	M_{ad}	6.92	7.57	8
	A_{ad}	18.82	5.78	0.52
	V_{ad}	59.14	71.36	75.70
	FC_{ad}	15.12	15.29	15.78
Element analysis (%)	C_{ad}	36.68	42.93	46.49
	H_{ad}	5.39	5.07	6.51
	O_{ad}	31.84	37.17	37.82
	N_{ad}	0.31	1.41	0.14
	S_{ad}	0.04	0.13	0.52
Heating value	Q_{net} (kJ/kg)	13398	14400	18875

4.1 Effect of Temperature

Fig.4 shows the variation of volume fraction of gas products generated by three kinds of biomass with the increasing of temperature based on 3800 calculated data The volume fractions of both H_2 and CH_4 increased sharply at the temperature from 600°C to 900°C. The mass flux rose from 0 to 72kg/h for H_2 and from 0 to 1000kg/h for CO. It is indicated that an increment in temperature is helpful to produce more H_2 and CO via absorbing heat reaction of methanation. However, as the temperature rose, the volume fraction was reduced by 65%-70% for CO_2, and was almost equal to zero for CH_4 over 800°C. The mass flux ranged from 1380kg/h to 500kg/h for CO_2 and from

300kg/h to 10kg/h for CH_4. It was in agreement with the results from the reference [6-8], It is mainly because CO_2 generated by the water and gas reaction and coke gasification reactions were consumed. Moreover, the methanation reaction was an exothermic reaction, the increment in temperature caused the generated CH_4 to transform into H_2 and CO, and hence leading to the drop in the volume fraction of CH_4. However, the CH_4 yield has surpassed the yield of H_2 for the rice husk gasification at less than 600°C, and was less than the yield of H_2 at the temperature over 600°C. What is more, the turning point is larger than 600°C for the corn straw and pine sawdust.

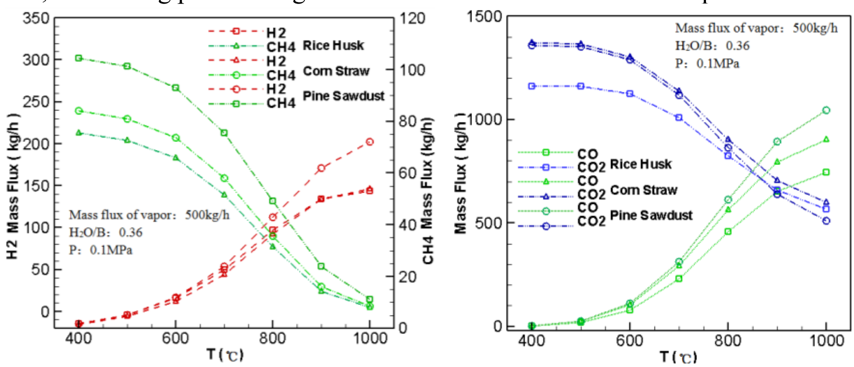


Fig. 4. Variations of volume fraction of gas products with the increasing of temperatures.

4.1. Effect of Pressure

Pressure is another key factor in the processing of biomass gasification based on 2600 calculated data. Fig.5 shows the variation of the gas products mass fluxes with the increasing of pressure at the mass flux of 500kg/h and the temperature of 800°C.

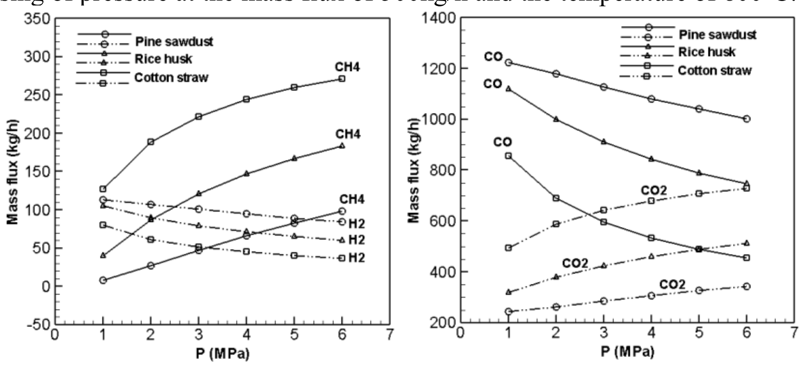


Fig. 5. Effect of pressure on mass fluxes of gas products.

As pressure rose from 1 MPa to 6 MPa, the mass flux of H_2 decreased from 113 kg/h to 84 kg/h for the pine sawdust, from 80 kg/h to 37 kg/h for the rice husk and from 105 kg/h to 60 kg/h for the corn straw. The mass fluxes of CO_2 rose from 243 kg/h to 343 kg/h for the pine sawdust, from 494 kg/h to 729 kg/h for the rice husk and from 319 kg/h to 513 kg/h for the corn straw. The mass flux of CO decreased from 1223 kg/h to 1002 kg/h for the pine sawdust, from 856 kg/h to 454 kg/h for the rice husk and from 1220 kg/h to 746 kg/h for the corn straw. The mass flux of CH_4 rose from 8 kg/h to 98 kg/h for the pine sawdust, from 127 kg/h to 271 kg/h for the rice husk and from

40 kg/h to 183 kg/h for the corn straw. It can be seen in Fig.5, as pressure rose, the mass flux of CH_4 rose, while the mass flux of both H_2 and CO reduced. It was in agreement with the results from the reference [9-10], in which the CO_2 production is encouraged at high temperature as a consequence of the occurrence of side-reactions. On the contrary, the CO_2 production was decreased as pressure ascended. Meanwhile, the pressure ascension drove the steam reforming reaction of the methane, causing the reduction in H_2 production. Similarly, the mass fluxes of gas products have the highest value for the pine sawdust, while the lowest value for the rice husk. It relied on the contents of C, H and H_2O in the biomass.

4.2. Effect of Various Gasify Agents

Here, H_2, CO_2, O_2, H_2O and H_2O-O_2 were used as the gasify agents, the rice husk was used as biomass that was sent into the gasifier at mass flux of 1400 kg/h at 700°C and 3Mpa. Fig.6 shows the volume fractions of gas at the ratio of the gasify agent to biomass masses (G/B) based on 7900 data.

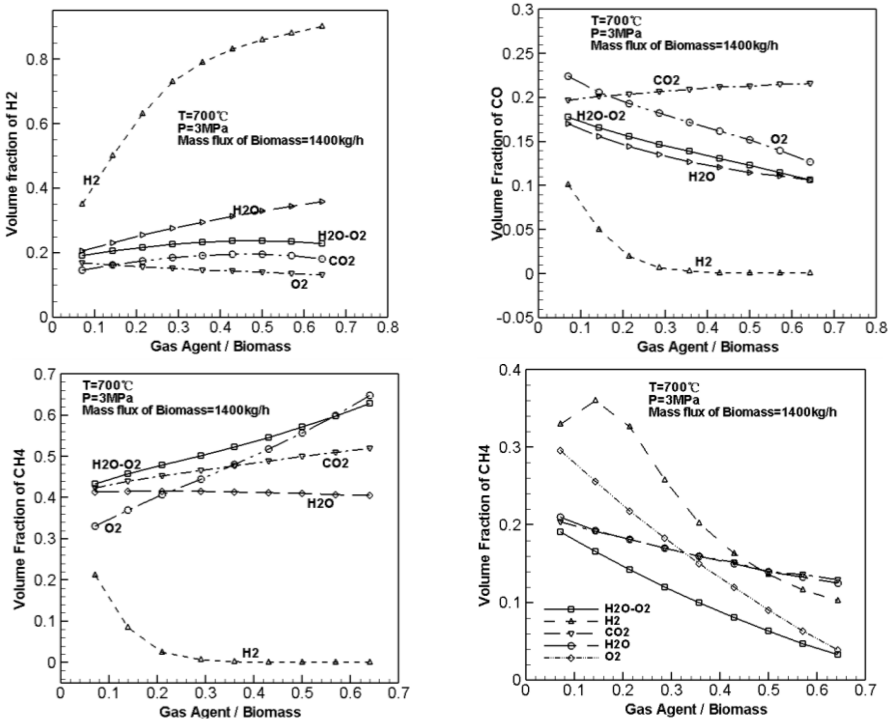


Fig. 6. Volume fractions of gas products at various G/B.

The volume fraction of H_2 was the highest in H_2 at various G/B. The volume fraction of H_2 reduced from 0.35 to 0.9 in H_2, ascended from 0.21 to 0.35 in H_2O, steadied in 0.19 to 0.2 in H_2O-O_2, dropped from 0.17 to 0.13 in O_2, and went up from 0.15 to 0.18 in CO_2. The volume fraction of CO was the highest in O_2 at G/B from 0.085 to 0.14 and in CO_2 over G/B of 0.14. The volume fraction of CO reduced from 0.225 to 0.125 in O_2, from 0.18 to 0.12 in H_2O-O_2, from 0.17 to 0.1 in H_2O and from 0.1 to 0 in H_2, while ascended from 0.19 to 0.21 in CO_2. The volume fraction of CO_2 was the highest in H_2O-O_2 at G/B from 0.22 to 0.56 and in O_2 over G/B of 0.56. The volume fraction of

CO_2 rose from 0.44 to 0.62 in H_2O-O_2, from 0.33 to 0.64 in O_2 and from 0.42 to 0.5 in CO_2, steadied in 0.42 in H_2O, while reduced from 0.22 to 0 in H_2. The volume fraction of CH_4 was the highest in CO_2 at less than G/B of 0.5 and in H_2O-O_2 over G/B of 0.5. The CH_4volume fraction rose from 0.205 to 0.1 in CO_2, went up from 0.33 to 0.36 and then dropped to 0.12 in H_2, and decreased from 0.19 to 0.04 in H_2O-O_2, from 0.3 to 0.03 in $O_2$and from 0.21 to 0.15 in H_2O.

5. Conclusions

Aspen Plus software was used to simulate the gasification processing of three kinds of biomasses at various gasification temperature and pressure when using various gasify agents. The gasification characteristics of herbaceous and woody plants were compared and the factors affecting gas were discussed. It comes to conclusions as follows.

1) The mass fluxes and volume fractions of gas products mainly depend on the contents of C, H and H_2O in the biomass. The highest mass flux and volume fractions of H_2, CH_4 and CO for pine sawdust gasification since it involves the highest mass fractions of C, H and H_2O.
2) Pressure has a reverse effect on the variations of volume fractions of gas products compared to temperature. The asension in pressure reduced the volume fractions of both H_2 and CO and increased the volume fractions of both CH_4 and CO_2. On the contrary, the increment in temperature surged the volume fractions of both H_2 and CO and decreased the volume fractions of both CH_4 and CO_2. The mass flux of gas reached the highest for pine sawdust.
3) It found that the gas consisted of CH_4 and H_2 if H_2/B is larger than 0.36.

References

[1] Valente A, Iribarren D, Dufour J. Life cycle sustainability assessment of hydrogen from biomass gasification: A comparison with conventional hydrogen. In. J. Hydro. Energy, 2019; (8):1-11.

[2] Fremaux S, Beheshti S, Ghassemi H, Shahsavan-Markadeh R. An experimental study on hydrogen-rich gas production via stam gasification of biomass in a research-scale fluidized bed. Energy conversion and management. 2015; (91):427-432.

[3] Wang TG, Sun L, Zhang XD, Li XF. The study of the behaviour of hydrogen released from biomass pyrolysis. Journal of Shandong University of Technology. 2006; 20(5):41-43.

[4] Vikram S, Rosha P, Kumar S, et al. Thermodynamic analysis and parametric optimization of steam-CO_2 based biomass gasification system using Aspen PLUS. Energy 2022; 241:1-9.

[5] Chen Q. Study and optimization of biomass gasification in a high-temperature entrained-flow gasifier for syngas. Doctor paper of Zhejiang University, 2012; 39-41.

[6] Alnouss A, Parthasarathy P, Shahbaz M, et al. Techno-economic and sensitivity analysis of coconut coir pith-biomass gasification using ASPEN PLUS. Appl. Energy 2020; 261(1); 1-10.

[7] Lv P, Yuan Z, Ma L, et al. Hydrogen-rich gas production from biomass air and oxygen/steam gasification in a downdraft gasifier. Renew. Energy 2007; 32(13): 2173-2185.

[8] Alnouss A, McKay G, Al-Ansari T. Enhancing waste to hydrogen production through biomass feedstock blending: A techno-economic-environmental evaluation. Appl. Energy 2020; 266(15); 1-14.

[9] Kitzler H, Pfeifer C, Hofbauer H. Pressurized gasification of woody biomass- variation of parameter. Fuel Process Technol 2011; 92(5): 908-914.

[10] Berrueco C, Recari J, Gǜell B M, Alamo G D. Pressurized gasification of torrefied woody biomass in a lab scale fluidized bed. Energy 2014; 70(1); 68-78.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230210

A Two-Dimensional Numerical Model for Surface Flows with Flexible Vegetation

Chen CHEN[a], Donghua ZHAO[a], Yao WU [a], Yanhao ZHENG [a], Yu DENG [a], Xiangxiang ZHANG [a], Min GAN [b,1] and Xijun LAI [b]

[a] *Shanghai Waterway Engineering Design and Consulting Co., Ltd., Shanghai 200120, China*

[b] *Key Laboratory of Watershed Geography Sciences, Nanjing Institute of Geography & Limnology, Chinese Academy of Sciences, Nanjing 200062, China*

Abstract. According to the demand of large-scale simulation, a numerical simulation model for vegetated flows was developed based on two-dimensional unsteady flow equations. According to the experimental flume results, the influence of aquatic plants on water flow was investigated. The vertical distribution characteristics of the drag coefficient induced by flexible vegetation at different positions of vegetation patches were obtained. The drag coefficients obtained from the experiments were substituted into the model to verify the model's reliability and explore the flow characteristics of open channels with vegetation patches. The results show that: 1) the simulation results of the model are in good agreement with the measured data, which can well simulate the water flow process under the influence of vegetation patches, especially for the far areas with weak three-dimensional effect; 2) The water-blocking effect of the vegetation patch reduces the flow velocity inside the patch and increases the flow velocity in the non-vegetated area on the side, forming a longitudinal shear effect at the edge. The proposed model can support the dynamic analysis of vegetation in rivers, lakes, and estuarine wetland systems.

Keywords. Vegetated flows, open channel, mathematical model, vegetation patches, flexible plant

1. Introduction

Aquatic plant is one of the essential components of the river and lake aquatic ecosystem. Most aquatic plants grow together in communities, forming vegetation zones. The vegetation belt formed by aquatic vegetation has a direct impact on the flow movement of rivers and lakes. Moreover, the changes in the movement structure and turbulence intensity of water flow also affect the circulation of material and energy in the water body. For example, the transport and settlement of sediment or nutrients and the exchange of materials at the water-soil interface can all be influenced [1–3]. The restoration of aquatic ecosystems is mainly based on the revival of aquatic vegetation. Although this method is widely used, the mechanism of ecological and environmental

[1] Min Gan, Key Laboratory of Watershed Geography Sciences, Nanjing Institute of Geography & Limnology, Chinese Academy of Sciences, 200062 Nanjing, China; E-mail: mgan@niglas.ac.cn.
This work was partly supported by the National Natural Science Foundation of China (Grant No.: 42171012).

effects of aquatic vegetation still needs to be further investigated. The impact of aquatic vegetation on water flow is the basis for revealing its ecological and environmental effects [4].

A large number of flume experiments and field observations have been carried out, which provided important data for understanding the impact of vegetation on water flow, sediment, and water quality [5,6]. To better assess and predict the influence of aquatic plants on water flow in the natural environment, the numerical simulation of water flow with vegetation has attracted wide attention. The researches of numerical simulation include the development of the numerical simulation method, revealing the process mechanism of vegetation affecting water flow, and explaining the effects of the plant on the water ecological environment, etc. [7–11]. The 3-D numerical method based on the turbulence model is the primary method for the numerical simulation of flow in open channels with vegetation. However, the two-dimensional model can satisfy the requirements of engineering planning design and its influence study because of its simplicity and practicability. Therefore, the two-dimensional model has been widely used in large-scale calculation and analysis of rivers, lakes, estuaries and wetlands. Due to the differences of plant morphology and characteristics, the magnitude of the influence of various vegetation on water flow movement is significantly different.

Potamogeton Malaianus is a widely distributed species in the rivers and lakes of China [12]. In this paper, based on the flume experiment results of flexible vegetation Potamogeton Malaianus, the additional resistance coefficient of vegetation is obtained. A numerical flow model of wetlands with flexible vegetation was established on the basis of the two-dimensional unsteady flow equations, which provides a scientific tool for understanding the influence of vegetation patches on water flow and dynamic analysis of large-scale river, lake and estuary wetland systems.

2. Materials and Methods

2.1. Numerical Models

- Governing equations

The flow state of the water body can be expressed by using the depth-averaged continuity and motion equations:

$$\frac{\partial \boldsymbol{U}}{\partial t}+\frac{\partial \boldsymbol{F}(\boldsymbol{U})}{\partial x}+\frac{\partial \boldsymbol{G}(\boldsymbol{U})}{\partial y}=\boldsymbol{B}(\boldsymbol{U}) \tag{1}$$

Where t is time; x and y represent the directions in the Cartesian coordinates; $\boldsymbol{U}$ is the vector of the state variables; $\boldsymbol{F}$ and $\boldsymbol{G}$ are the flux vectors along the x and y directions, respectively; $\boldsymbol{B}$ is the vector of the source term. The terms $\boldsymbol{U}$, $\boldsymbol{F}$, $\boldsymbol{G}$, and $\boldsymbol{B}$ have the following expressions:

$$\boldsymbol{U}=(h,hu,hv)^T \tag{2}$$

$$\boldsymbol{F}=(hu,hu^2+\frac{1}{2}gh^2,huv)^T \tag{3}$$

$$\boldsymbol{G} = (hv, huv, hv^2 + \frac{1}{2}gh^2)^T \tag{4}$$

$$\boldsymbol{B} = \left(0, gh\left(S_{0x} - S_{fx}\right) + S_{veg,x}, gh\left(S_{0y} - S_{fy}\right) + S_{veg,y}\right)^T \tag{5}$$

where h is the water depth; u and v are the depth-averaged velocity along the x and y directions, respectively; T is the symbol of transpose; g is the gravitational acceleration. S_{0x} (S_{fx}) and S_{0y} (S_{fy}) are the bottom (drag) slope terms along the x and y directions, respectively. $S_{veg,x}$ and $S_{veg,y}$ are the additional resistance forces from the vegetables along the x and y directions and have the following expressions:

$$S_{veg,x} = \frac{1}{2}C_p au|u| \tag{6}$$

$$S_{veg,y} = \frac{1}{2}C_p av|v| \tag{7}$$

where C_p is the drag coefficient of the vegetables; a is the vegetable density. C_p is calculated as follows:

$$C_p = \frac{2(gJ - \overline{\Delta u'w'}/\Delta z)}{a(z)u^2} \tag{8}$$

Where J is the hydraulic gradient; $u'w'$ is the Reynolds stress; z is the vertical direction in the Cartesian coordinates; Δz is the vertical distance; $a(z)$ is the vegetable density at z location.

- Numerical solution

In this study, the unstructured grid finite volume method is used to solve the two-dimensional unsteady flow equation [13]. The HLLC (Harten-Lax-van Leer-contact) scheme with high spatial resolution was used to approximate the interface flux. The equation (1) is integrated on any given control volume element, and the Gauss Green formula is used to convert the area integral into a line integral. The original two-dimensional normal flux solution problem is then converted into solving the Riemann problem in a one-dimensional local coordinate system by using the rotational invariance of the convection flux of the control equation [14]. Therefore, only the flux of the normal velocity needs to be calculated to improve the calculation efficiency.

$$\frac{d\boldsymbol{U}}{dt} = -\frac{1}{A}\sum_{j=1}^{m} \boldsymbol{T}(\theta)^{-1}\boldsymbol{F}_j(\overline{\boldsymbol{U}})\, L_j + \overline{\boldsymbol{B}(\overline{\boldsymbol{U}})} \tag{9}$$

where A is the area of the control volume; j is the edge index of the control volume; m is the total number of edges; L_j is the length of j^{th} edge; θ is the intersection angle between x axis and the norm vector (anticlockwise direction of x axis); $\boldsymbol{T}(\theta)$ is the transformation matrix of θ; $\boldsymbol{B}(\overline{\boldsymbol{U}})$ is the comprehensive source terms including the additional vegetation resistance. For more details of the HLLC solution, the readers can refer to reference [14].

- Definite solution condition

To solve the aforementioned equations in a specific condition, the corresponding initial and boundary conditions need to be given simultaneously. The initial conditions are given as follows:

$$\boldsymbol{U}_0 = \left(h|_{t=\boldsymbol{t}_0}, hu|_{t=\boldsymbol{t}_0}, hv|_{t=\boldsymbol{t}_0}\right)^T \tag{10}$$

where the subscript t_0 represent the variable values at the initial condition.

The boundary conditions can be divided into two types which are land boundary and open boundary, respectively. For a given water level (depth) condition, $h(t)|_\Gamma$ is known. According to the known boundary condition on the left-hand state, the unknown boundary condition on the right-hand state can be calculated by the characteristic relation:

$$u_R = u_L + 2\left(\sqrt{gh_L} - \sqrt{gh_R}\right) \tag{11}$$

where subscript R and L represent the variables at the left-hand and the right-hand sides, respectively.

For the land boundary condition, $h_R = h_L$, $u_R = -u_L$, and $v_L = v_R$. The normal flux can be calculated based on the left and right state conditions.

2.2. Physical Experiments

The results from a variable slope flume were used for the validation of the model proposed in this study. As is shown in **Figure 1**, the length and the width of the flume are 38.00 m and 0.80 m, respectively. The vegetation belt is located 10.00 m away from the front of the flume. Its length is 1.60 m and its width is 0.40 m. The height of the vegetation canopy is about 0.14 m. More details of the flume experiments can be found in reference [15].

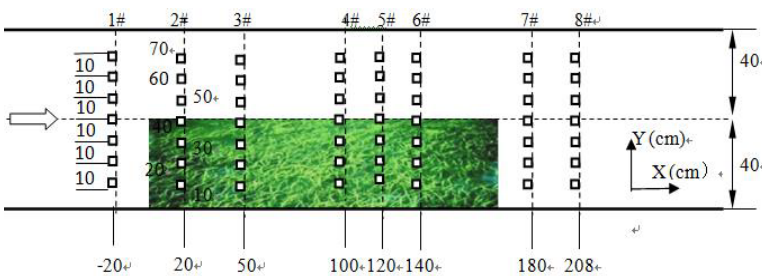


Figure 1. Location of the vegetable patch and the distributions of the observation sections.

3. Results & Discussion

According to the experimental results, the vegetation resistance coefficient can be calculated from the measured U_i, $-\overline{u'w_i'}$, and the hydraulic gradient. The vertical distribution of C_p at different cross sections is shown in **Figure 2**. In general, the resistance coefficient above the vegetation canopy is between 1.4-4.0. When the relative water depth is less than 0.27, the drag coefficient decreases with the increase of water depth. However, when the relative water depth is greater than 0.80, the drag coefficient increases with the increase of water depth. The vertical distribution of the additional resistance coefficient varies greatly due to the swing of aquatic plants. Therefore, the vertical average values were used in the two-dimensional model of this study.

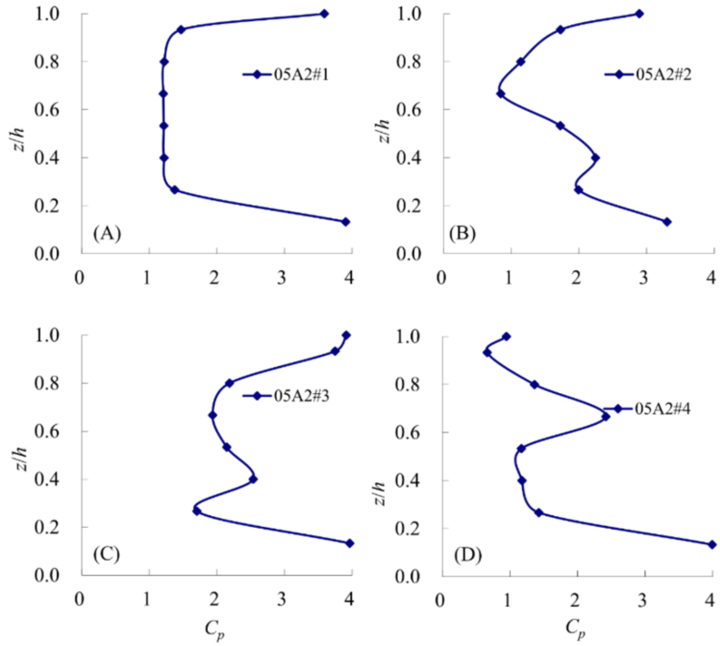


Figure 2. The vertical distribution of C_p at different cross sections.

Based on the size information of the flume, the mesh size of the two-dimensional model was specified as 0.025 m×0.100 m. The flow velocity at the water inlet was set as 0.5 cm/s. The water level at the inlet was set as 0.30 m which is the same as the actual water depth under the experimental condition. The slope of the flume is 0.005, while the manning roughness coefficient is 0.016.

Taking sections 2#, 3#, 4# and 6# as examples, the simulated current velocity was compared with the measurements. Seven average velocities can be obtained for one cross-section when the vertical velocity is averaged (**Figure 1**). **Figure 3** shows the comparison of the model results and the measurements. Overall, the model results get closer to the measurements along the water flow direction. When the water just enters the vegetation area, its three-dimensional flow characteristics are significant because of the mutual influence of the water flow and the vegetation patch. With the downward flow of the water, the flow movement is adjusted to be more stable, and its two-dimensional flow characteristics become more prominent. **Figure 3** illustrates that the flow with flexible vegetation can be simulated by increasing the additional resistance coefficient, and the model developed in this study can be used to analyze the water flow with a weak three-dimensional effect.

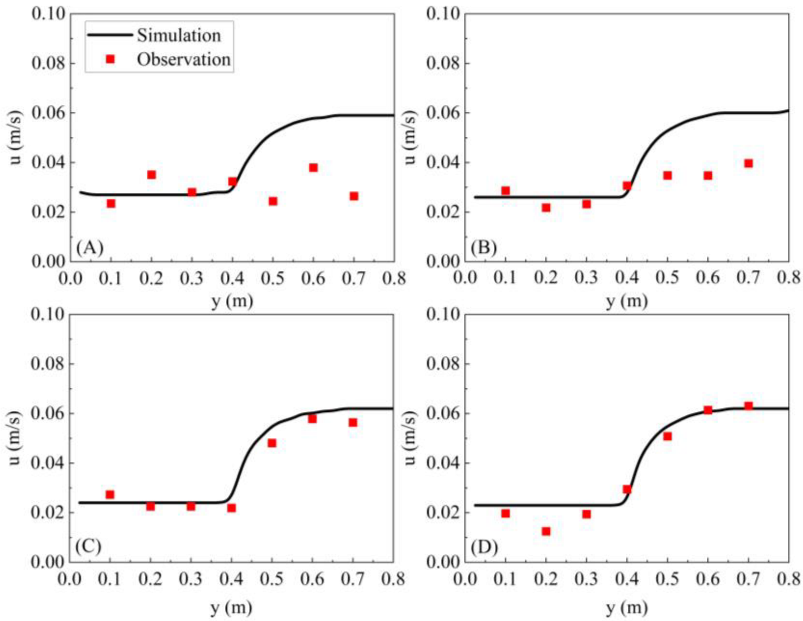


Figure 3. Comparison of the observed and simulated current velocities at sections 2# (A), 3# (B), 4# (C), and 6# (D).

The streamline and the current velocity field from the model results are shown in **Figure 4**. At 0.20 m upstream of the vegetation patch (#2 cross-section), the phenomenon of the flow around the vegetation patch appears due to the water-blocking effect. The flow velocity at the vegetation and non-vegetation areas are weakened and enhanced, respectively. In addition, the flow velocity in the vegetation area decreases along its flow direction because of the water-blocking effect. However, the flow velocity becomes larger when the non-vegetation site is farther away from the vegetation area. The flow velocity is still at a low-velocity state in the areas near the downstream of the vegetation area, indicating that the flow velocity needs to flow some distances to recover to the state without the influence of the vegetation patch.

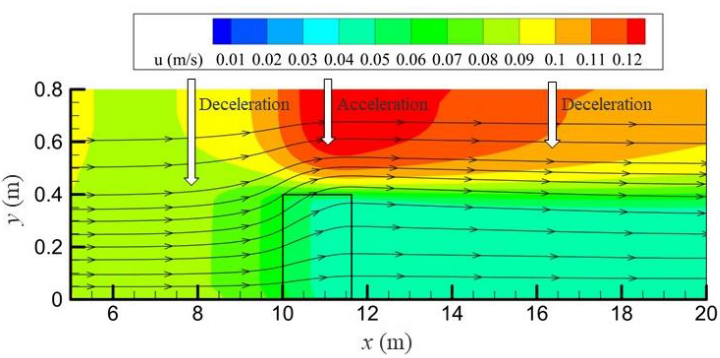


Figure 4. The flow line and current velocity field from the model results.

4. Conclusions

In this paper, a two-dimensional unsteady flow model with the consideration of vegetation is established. The model is validated and calculated by considering the influence of vegetation patches on open channel flow based on the experimental flume results. The common aquatic plant *Potamogeton Malaianus* in the river and lake waters of China was used as the object to obtain the experimental data.

The results show that: 1) The model simulation results are in good agreement with the measured data and can simulate the flow process influenced by vegetation patches, especially in the remote area where the three-dimensional effect is weak; 2) The water-blocking action of the vegetation patches reduces the flow velocity inside the vegetation patches and increases the flow velocity in the vegetation-free areas, resulting in longitudinal shear at the edges.

References

[1] Nepf H, Koch E. Vertical secondary flows in submersed plant-like arrays. Limnology and Oceanography. 1999; 44(4): 1072-1080.
[2] Chen Y, Kao S. Velocity distribution in open channels with submerged aquatic plant. Hydrological Processes. 2011; 25(13): 2009-2017.
[3] Gambi M, Nowell A, Jumars P. Flume observations on flow dynamics in zostera marina (eelgrass) beds. Marine Ecology Progress Series. 1990; 61(1/2): 159-169.
[4] Fagherazzi S, Kirwan M, Mudd S, Guntenspergen G, Temmerman S. Numerical models of salt marsh evolution: ecological, geomorphic, and climatic factors. Reviews of Geophysics. 2012; 50(1): RG1002.
[5] Yamasaki T, Jiang B, Janzen J, Nepf H. Feedback between vegetation, flow, and deposition: a study of artificial vegetation patch development. Journal of Hydrology. 2021; 598: 126232.
[6] Zhang Y, Lai X, Ma J, Zhang Q, Yu R, Yao X, Deng H. Field study on flow structures within aquatic vegetation under combined currents and small - scale waves. Hydrological Processes. 2021; 35(4): e14121.
[7] Li C, Xie J. Numerical modeling of free surface flow over submerged and highly flexible vegetation. Advances in Water Resources. 2011; 34(4): 468-477.
[8] Lu J, Dai H. Numerical modeling of pollution transport in flexible vegetation. Applied Mathematical Modelling. 2018; 64: 93-105.
[9] Stoesser T, Wilson C, Bates P, Dittrich A. Application of a 3D numerical model to a river with vegetated floodplains. Journal of Hydroinformatics. 2003; 5(2): 99-112.
[10] Huai W, Zhang J, Katul G, Cheng Y, Tang X, Wang W. The structure of turbulent flow through submerged flexible vegetation. Journal of Hydrodynamics. 2019; 31(2): 274-292.
[11] Pu J, Ricardo A, Ferreira R. A Review on hydrodynamics of free surface flows in emergent vegetated channels. Water. 2020; 12(4): 1218.
[12] Dong B, Zhou Y, Jeppesen E, Qin B, Shi K. Six decades of field observations reveal how anthropogenic pressure changes the coverage and community of submerged aquatic vegetation in a eutrophic lake. Science of The Total Environment. 2022; 842: 156878.
[13] Zhao D, Shen H, Tabios G, Lai J, Tan W. Finite - volume two - dimensional unsteady - flow model for river basins. Journal of Hydraulic Engineering. 1994; 120 (7): 863-883.
[14] Lai X, Jiang J, Liang Q, Huang Q. Large-scale hydrodynamic modeling of the middle Yangtze river basin with complex river–lake interactions. Journal of Hydrology. 2013; 492: 228-243.
[15] Zhang Y, Lai X, Jiang J. The impact of plant morphology on flow structure: comparative analysis of two types of submerged flexible macrophyte. Hydrological Sciences Journal. 2016; 61(12): 2226-2236.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230211

Neural Network Prediction of Dynamic Characteristics of Soft Soil Disturbance

Minggang PENG[1] and Jinjin SUN
China Power Construction Railway Construction Investment Group Co., Ltd., Beijing 100070, China

Abstract. In recent years, many large-scale projects have been built on thick soft soil foundations. Due to insufficient understanding of the engineering properties of soft soil, many major engineering accidents have occurred. Based on the intersection project of Nangangdao Bridge and subway tunnel in Fuzhou, in order to study the disturbance degree of pile foundation construction to surrounding soil under the influence of multi factor coupling, direct shear test under different disturbance factors (vertical pressure, consolidation time, shear rate, disturbance degree) were carried out on soft soil in Fuzhou, revealing the shear strength characteristics of soft soil in Fuzhou and the sensitivity of peak shear stress to various disturbance factors. The experimental results indicate that all disturbance factors have a significant impact on the stress-strain curve of the soil. As the shear rate increases, the stress-strain curve gradually transitions from strain hardening to strain softening. Propose a sensitivity analysis method to normalize the peak shear stress under various disturbance factors, and obtain the maximum fitting slope of peak shear stress to consolidation time, the highest sensitivity, followed by shear rate, and the minimum disturbance degree. A neural network prediction model for the peak shear strength of saturated soft clay was established using the principle of artificial neural networks. This model has the characteristics of considering multiple factors, high fitting degree, and strong predictive ability.

Keywords. Artificial neural network, dynamic characteristics, degree of disturbance, direct shear test, peak shear stress, saturated soft clay

1. Introduction

With the continuous advancement of urbanization in China, the scale of infrastructure construction in the southeast coastal areas of China is increasing, and a large amount of muddy clay is distributed in the coastal areas. The strength, deformation characteristics, and treatment methods of saturated soft clay have become the primary issues faced in the construction process of coastal areas. This type of soft clay has characteristics such as high water content, high compressibility, large void ratio, low shear strength, and significant rheological deformation. Therefore, soft soil layers in natural state should not be directly used as foundation in infrastructure projects [1-2].Therefore, it is particularly important to understand the strength and deformation characteristics of soft soil layers under different natural conditions. In order to meet the requirements for the strength and deformation of foundation and subgrade, it is also possible to reinforce natural soft soil layers [3-5]. In the process of pile foundation construction, various

[1] Minggang Peng, China Power Construction Railway Construction Investment Group Co., Ltd., Beijing 100070, China; E-mail: 478804347@qq.com.

factors have a significant impact on the shear strength of soft soil foundation. There is relatively little research on the achievements in this area, so it is necessary to study the relationship between the disturbance conditions during the construction process and the changes in soft soil strength.

In the research of geotechnical constitutive relations, there are more and more models with complex expression forms, but there are some drawbacks in the application in practical engineering. In practical engineering, engineers mainly focus on the actual response characteristics and practicality of working conditions, rather than the parameters and expression forms of constitutive models. Artificial neural networks (ANNs) have strong advantages in dealing with such problems, and incorporate practical forces into the study of constitutive models in geotechnical engineering [8-9].

Among the forms of soil failure, there are soil cracking due to excessive tensile force, as well as damage due to excessive shear force. In most engineering problems, such as soil slope stability issues, foundation bearing capacity issues, soil disturbance damage issues, foundation pit excavation soil deformation and safety and stability issues, shear strength of soil is mainly considered. There are many factors influencing the shear strength of soil, and they are also very complex. According to the sensitivity analysis method proposed by Mohammadizadeh, SM [10], it is of great significance to study the sensitivity of various influencing factors to the strength characteristics of saturated soft clay, which can provide important theoretical support for geotechnical engineering design and safety and stability prediction analysis.

2. Sample Preparation and Testing Details

2.1. Description of the Tested Soil

The soil for the test was taken from the middle section of the bottom plate elevation of the subway tunnel roof at the intersection of the Nangang Road Bridge on Fuzhou New City Expressway and the Metro Line 6 tunnel, with a depth of 5-19m. A thin-walled soil sampler was used to penetrate the soil sample, and the soil sample grade was Grade I. After sampling, all joints were sealed with adhesive tape, wrapped with plastic wrap to prevent the evaporation of the original soil moisture, and packaged with wooden boxes for centralized shockproof packaging. The soil sample belongs to the typical marine sedimentary soft soil in Fuzhou area.

Basic physical tests shall be conducted to determine the natural weight of the undisturbed soil. The alcohol combustion method shall be used to repeatedly determine the natural water content of the undisturbed soil samples. The specific gravity of the soil particles shall be determined using the pyknometer method, and the limit water content shall be determined using the liquid plastic limit combined measurement method. The basic physical indicators are shown in Tab. 1 below.

Table 1. Basic Physical Indexes of Undisturbed Soil.

Natural soil weight γ/kN*m^{-3}	Natural water content w/%	Proportion Gs	Liquid limit w_L/%	Plastic limit w_P/%	Plasticity index I_P
17.97	33.6	2.70	65.6	24.6	41.0

The undisturbed soil sample used in the test is a fine grained soil that is not easily permeable to water. Using the vacuum saturation method, the undisturbed soil sample with a ring knife is placed in a saturator, and then placed in a vacuum cylinder for air extraction saturation. The standing time is set to 12 hours to fully saturate the undisturbed soil sample.

The remolded soil sample used in the test is the undisturbed soil sample that has been fully air dried, removed impurities, hammered and crushed through a 0.45mm geotechnical sieve, and repeatedly measured the moisture content after air drying. A remolded soil sample with the same dry density and moisture content is prepared using the sample pressing method based on the volume of the ring knife and the dry density and moisture content of the undisturbed soil, and then the remolded soil sample is fully saturated.

The artificially prepared structural soil samples with different disturbance levels used in the test are prepared from 1% and 2% cement soil samples, using remolded soil that has passed a 0.45mm geotechnical sieve, cement with a model of 525 Portland, edible salt, and water as materials. Adding cement to the remolded soil can establish a bond between soil particles, and the dissolved edible salt in the remolded soil can construct the macropore characteristics of soft clay, The specific preparation method is as follows:

(1) After fully mixing and mixing the mixture described above, place it in a ring knife, compact it using a sample pressing method, add an appropriate amount of water, and place it in a humid environment to achieve initial setting.

(2) Place the soil sample in flowing water for 1 day, dissolve all the edible salt particles and take them away with the water flow.

(3) Saturate the sample by air extraction, and cure all samples for 3 days before conducting the test.

The proportion of manually prepared structural soil is shown in Table 2.

Table 2. Proportions of Manually Prepared Structural Soil.

Artificial structural soil	S_m/%	C_m/%	Edible salt/g
$C_{1.0}$	99	1	8
$C_{2.0}$	98	2	8

Note: S_m is the ratio of clay mass to the total mass of clay and cement; C_m is the ratio of the mass of cement to the total mass of clay and cement.

2.2. Test Plan

There are four groups of undisturbed soil samples and remolded soil samples each,the test plan is shown in the table below.

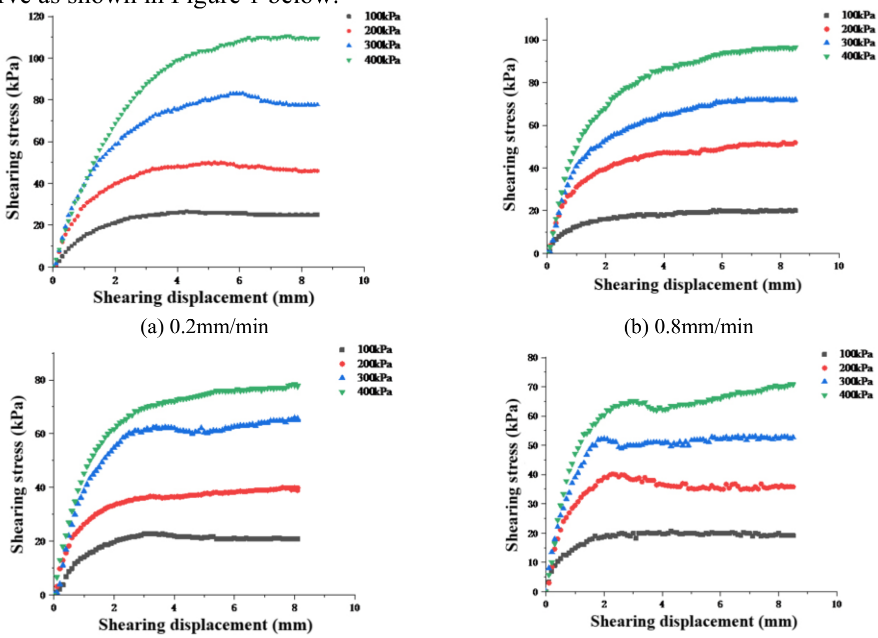
Tab.3 Direct Shear Test Plan

	Degree of sample disturbance	Shear rate (mm/min)	Consolidation time (min)	Vertical pressure (kPa)
Undisturbed sample	Undisturbed	0.2, 0.8, 1.6, 2.4	100	100, 200, 300, 400
Disturbed sample	Remodeling	0.2, 0.8, 1.6, 2.4	100	100, 200, 300, 400
	Reconsolidation	0.8	5, 40, 100	100, 200, 300, 400
	Artificial structural soil (1%, 2% cement soil)	0.8	100	100, 200, 300, 400

3. Analysis and Discussion of Test Results

3.1. Strength Characteristics of Undisturbed Specimens

For undisturbed soil samples with basically the same physical properties, measure the shear displacement and shear stress at the shear rates of 0.2mm/min, 0.8mm/min, 1.6mm/min, and 2.4mm/min, and draw the shear displacement shear stress relationship curve as shown in Figure 1 below.



(a) 0.2mm/min (b) 0.8mm/min

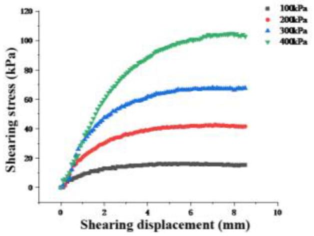
(c) 1.6mm/min (d) 2.4mm/min

Figure 1. Shear displacement shear stress curve of undisturbed soil.

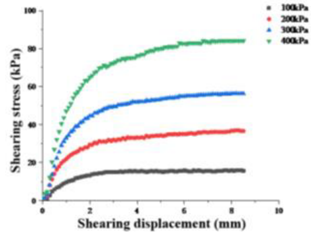
3.2. Strength Attenuation Characteristics of Disturbed Soil

- Strength characteristics of remolded soil

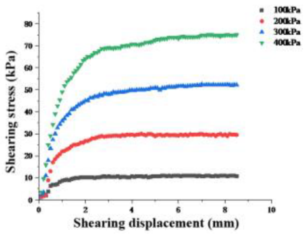
For remolded soil samples with basically the same physical properties, measure the shear displacement and shear stress at the shear rates of 0.2mm/min, 0.8mm/min, 1.6mm/min, and 2.4mm/min, and draw the shear displacement shear stress relationship curve as shown in Figure 2 below.



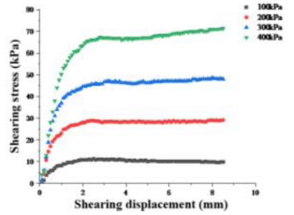
(a) 0.2mm/min



(b) 0.8mm/min



(c) 1.6mm/min

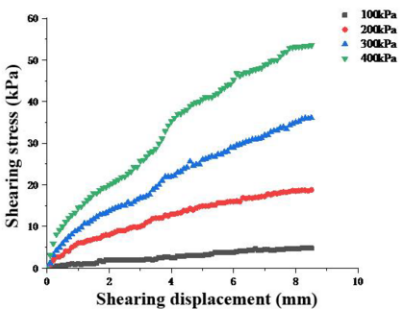


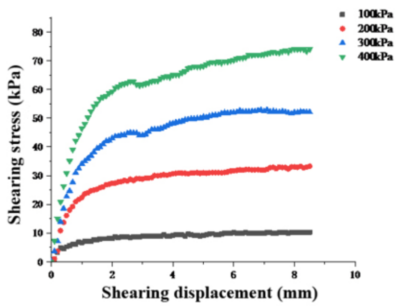
(d) 2.4mm/min

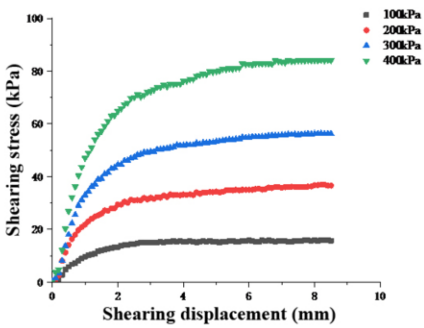
Figure 2. Shear displacement shear stress curve of remolded soil.

- Strength characteristics of reconsolidated soil

For remolded soil samples with basically the same physical properties, measure the shear displacement and shear stress at a shear rate of 0.8mm/min after the consolidation time of 5min, 40min, and 100min, and draw the shear displacement shear stress relationship curve as shown in Figure 3 below.







(a) 5min (b) 40min (c) 100min

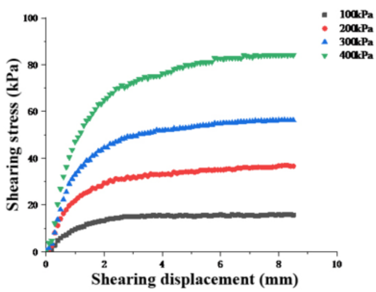
Figure 3. Shear displacement shear stress curve of remolded soil at different consolidation times.

- Strength characteristics of soils with different degrees of disturbance

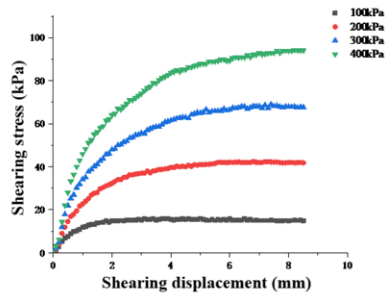
For artificial structural soil samples with basically the same physical properties and cement parameters of 1% and 2%, respectively measure the shear displacement and shear stress at a shear rate of 0.8mm/min, and draw the shear displacement shear stress relationship curve as shown in Figure 4 below.

The test results show that vertical pressure, shear rate, consolidation time (degree of consolidation), and degree of disturbance all affect the shear displacement and shear stress curve of soil. The higher the vertical pressure, the greater the peak shear stress; the higher the shear rate, the smaller the peak shear stress; the greater the peak shear stress; the greater the soil disturbance, the smaller the peak shear stress.

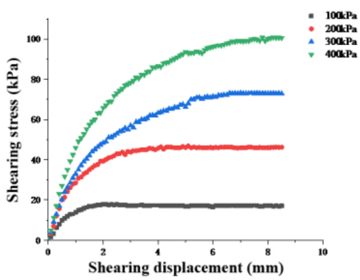
The degree of consolidation has the most significant effect on the peak shear stress in the shear displacement and shear stress curves. The shear displacement and shear stress curves of undisturbed and remolded soils gradually show a trend of strain softening with the increase of shear rate.



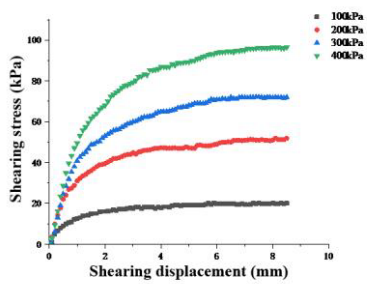
(a)Remolded soil



(b)1% cement soil



(c)2% cement soil



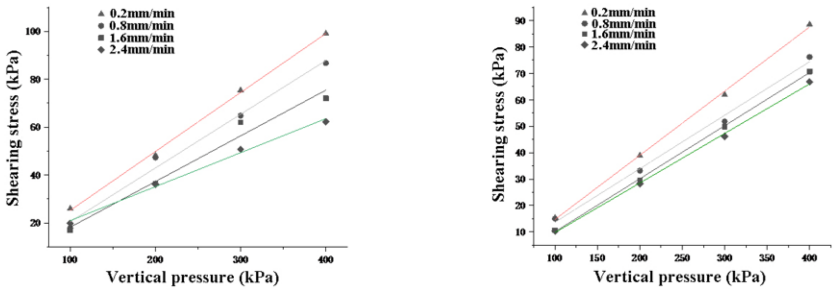
(d)Undisturbed soil

Figure 4. shear displacement shear stress curves of soils with different disturbances.

3.3. Study on the Influence of Various Factors on the Peak Shear Stress

- Effect of shear rate on peak shear stress

According to the shear displacement and shear stress test curves of undisturbed soil and remolded soil at the shear rates of 0.2mm/min, 0.8mm/min, 1.6mm/min, and 2.4mm/min, respectively, take the shear displacement Δ the shear stress corresponding to L=4mm is taken as the shear strength S and is the peak shear stress. The relationship between the peak shear stress and vertical pressure for undisturbed soil and remolded soil at different shear rates is shown in Figure 5 and 6.



(a) Remolded soil (b) Undisturbed soil

Figure 5. Relationship curve between peak shear stress and vertical pressure under different shear rates.

As can be seen from Figure 5, with the increase of shear rate, the peak shear stress under each vertical pressure presents a decreasing trend. When the vertical pressure of undisturbed soil and remolded soil is lower than that of 100kPa or 200kPa, the difference between the peak shear stress under each shear rate is small. When the vertical pressure of 100kPa is applied to undisturbed soil, the shear rate is large but the peak shear stress is also large. This phenomenon indicates that with the increase of vertical pressure, The effect of shear rate on peak shear stress is more obvious.

- Effect of consolidation time on peak shear stress

According to the direct shear test of remolded soil at a shear rate of 0.8mm/min under a consolidation time of 5min, 40min, and 100min, the relationship curve between the peak shear stress and vertical pressure is shown in Figure 6 below.

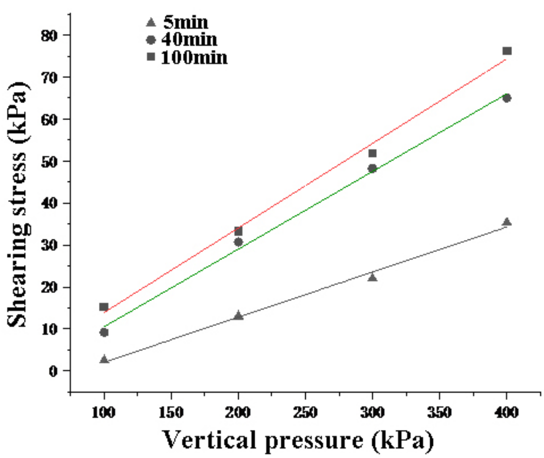


Figure 6. Relationship curve between peak shear stress and vertical pressure of remolded soil at different consolidation times.

As can be seen from Figure 6, with the increase of vertical pressure, the peak shear stress increases linearly. Under the same vertical pressure, the longer the consolidation time, the greater the peak shear stress. The peak shear stress increases most significantly during the consolidation time of 5 to 40 minutes, and the peak shear stress increases slowly during the consolidation time of 40 to 100 minutes. This phenomenon occurs because the main consolidation of soft clay is mostly completed within 40 to 100 minutes, With the discharge of pore water from soft clay, the effective stress of the soil gradually increases, and the peak shear stress increases significantly. After the primary consolidation of soft clay is completed, it enters the secondary consolidation stage, where pore water almost no longer discharges, soil particles undergo creep and adjustment, and the effective stress between soil particles increases slowly, and the peak shear stress slightly increases during the consolidation time of 40 to 100 minutes; The slope of the fitting curve increases as the consolidation time increases within the range of 5 to 100 minutes.

- The effect of disturbance on the peak shear stress

According to the prepared artificial structural soil samples with different cement content, direct shear tests of the soil were conducted at the same consolidation time and shear rate. The relationship curve between the peak shear stress and vertical pressure is shown in Figure 7 below.

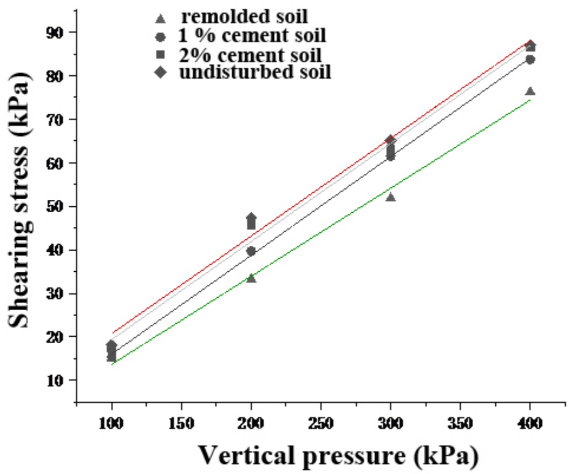


Figure 7. Shear Stress Peak and Vertical Pressure Relationship Curves with Different Disturbances.

As can be seen from Figure 7, under the same vertical pressure, the greater the degree of disturbance, the smaller the peak shear stress. The order of the peak shear stress is: undisturbed soil, 2% cement-soil, 1% cement-soil, remolded soil. Artificial structured soil has a better effect in simulating soft clay with different degrees of disturbance.

3.4. Sensitivity Discussion of Various Factors Affecting Peak Shear Stress

According to the research in section 2.3, the magnitude of peak shear stress is mainly related to shear rate, consolidation time, and degree of disturbance. In the range of shear rate from 0.2 to 2.4 mm/min, the reference is 2.4 mm/min, the reference is 100 min for the range of consolidation time from 5 to 100 min, and the reference is undisturbed soil; The variation degree of peak shear stress under various influencing factors is shown in the table below.

Table 4. Percentage of peak shear stress of undisturbed soil under various vertical pressures based on a shear rate of 2.4mm/min.

Shear rate /mmmin^{-1}	Percentage /%	Percentage at different vertical pressures /%			
		100kPa	200kPa	300kPa	400kPa
0.2	8	131	133	155	160
0.8	33	92	130	131	139
1.6	67	89	101	128	117
2.4	100	100	100	100	100

Table 5 Percentage of peak shear stress of remolded soil under various vertical pressures based on a shear rate of 2.4mm/min.

Shear rate /mmmin^{-1}	Percentage /%	Percentage at different vertical pressures /%			
		100kPa	200kPa	300kPa	400kPa
0.2	8	149	140	137	142
0.8	33	147	116	113	116
1.6	67	102	104	109	106
2.4	100	100	100	100	100

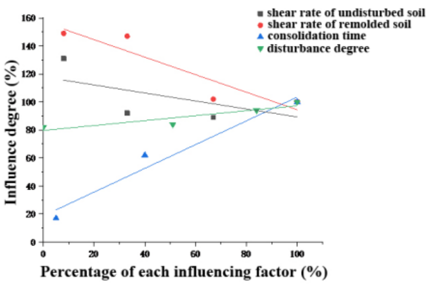
Table 6. Percentage of peak shear stress and vertical pressure of remolded soil based on consolidation time of 100min.

Consolidation time /min	Percentage /%	Percentage at different vertical pressures /%			
		100kPa	200kPa	300kPa	400kPa
5	5	17	32	42	47
40	40	62	93	93	85
100	100	100	100	100	100

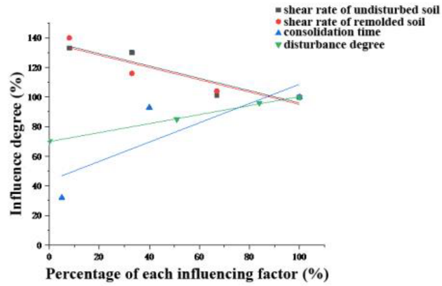
Table 7. Percentage of peak shear stress per vertical pressure based on different perturbation degrees.

Disturbance degree	Percentage /%	Percentage at different vertical pressures /%			
		100kPa	200kPa	300kPa	400kPa
Remolded soil	0	82	70	80	89
1% cement soil	51	84	85	95	97
2% cement soil	84	94	96	98	99
Undisturbed soil	100	100	100	100	100

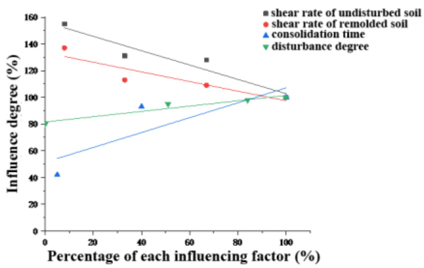
In the table above, the influence degree of each influencing factor on the peak shear stress varies. In order to explore the sensitivity of the magnitude of the peak shear stress to the shear rate, consolidation time, and degree of disturbance, the percentage of each influencing factor and the influence degree curve of the peak shear stress are calculated at 100kPa, 200kPa, 300kPa, and 400kPa, respectively:



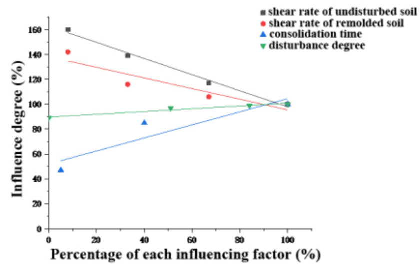
(a) 100kPa



(b) 200kPa



(a) 300kPa



(b) 400kPa

Figure 8. Curve of percentage of each influencing factor and influence degree of peak shear stress.

From the above Figure 8, it can be seen that the absolute value of the slope of the fitting line for consolidation time is the highest, followed by the absolute value of the slope of the fitting line for shear rate, and the absolute value of the disturbance fitting line is the lowest. This indicates that in direct shear tests of soil, the sensitivity of peak shear stress to consolidation time is the highest, followed by shear rate, and the sensitivity of disturbance is the lowest.

As the consolidation time increases, the degree of consolidation of soft soil increases, and the soil particles change significantly. The arrangement and distribution of particles gradually become tight, and the pores are gradually filled with broken small particles. The effective stress between soil particles gradually increases, leading to a gradual increase in soil strength. Compared to the consolidation time, the degree of disturbance in the soil is less sensitive to it. In the soil disturbance process described in the article, soil samples with different cement dosages indicate different particle arrangements and cementation effects. The sensitivity analysis results indicate that the shear resistance of soil is most significantly affected by effective stress.

4. 3 Neural Network Prediction Model for Strength of Saturated Soft Clay

In order to study the feasibility of using neural networks to establish the stress and strain of saturated soft clay and the impact of input variables and training samples on the performance of the model, different combinations of input variables and training samples were selected to establish a neural network prediction model for the peak shear stress of soft clay, namely Model A. Model A establishes a neural network model for nonlinear stress-strain curves of soft soil based on the results of direct shear tests under conditions such as multiple degrees of disturbance, multiple vertical stresses, and multiple shear rates for the same soil, peak shear strength τ as an output variable (Figure 9-a). The neural network model is expressed as follows:

$$\tau = f\left(\sigma_v, t, v, D\right)$$

Where, σ_v is the vertical stress, t is the consolidation time, v is the shear rate, and D is the degree of disturbance.

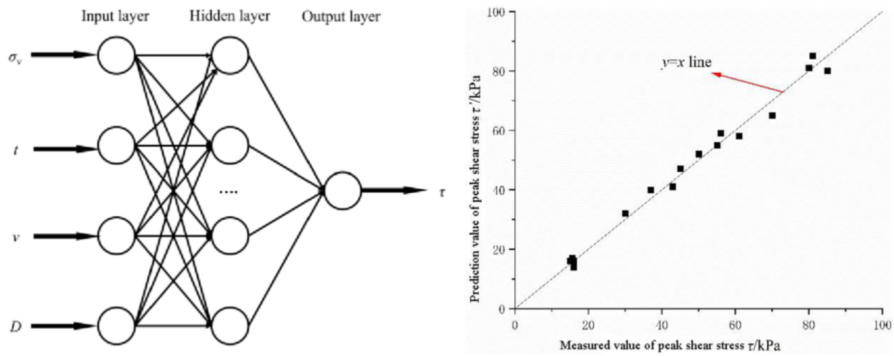


Figure 9. a) Structure of Soft Soil Neural Network Constitutive Model; b) Comparison of predicted and measured peak shear stresses in saturated soft clay under different disturbance levels

Using Matlab's neural network toolbox for programming operations. with σ_v,t,v,D are used as the inputs of the network, and shear stress is used as the output of the network. A neural network model is established. The structure of the BP neural network model is 5-13-1. The network is trained using the results of direct shear tests. Taking the peak shear stress of soft soil under different disturbance levels as an example, the predicted value and the measured value are shown in Figure 9-b. The two values are uniformly distributed on both sides of the y=x line, and the values are basically consistent.

5. Conclusion

In this paper, the undisturbed soil taken from the intersection of Fuzhou Nangangdao Bridge and subway tunnel is taken as the test object, and the strength and deformation characteristics of saturated soft clay under different disturbance factors (vertical stress, shear rate, disturbance degree, consolidation time) are analyzed by using strain controlled direct shear apparatus. A prediction model for peak shear stress under different disturbance factors was established based on artificial neural networks.

(1) The various disturbance factors in this article have varying degrees of influence on the shear displacement and shear stress curves of saturated soft clay, with consolidation time having the most significant impact on the curves. The shear displacement and shear stress curves of undisturbed and remolded soils gradually change from strain hardening to strain softening with the increase of shear rate.

(2) The concept of sensitivity analysis was proposed, and the sensitivity of peak shear stress to different disturbance factors was compared. Its sensitivity to consolidation time was the highest, followed by shear rate and disturbance degree.

(3) Based on artificial neural networks, a prediction model for peak shear stress was established, which can effectively predict and analyze the peak shear stress under the influence of various factors in the text.

Reference

[1] Xu AJ, Li YL. Construction control of deep pit engineering in Shanghai soft-soil areas. Journal of Geotechnical Engineering. 2006:1395-7.

[2] Sun DA, Shen HE. Experimental study on rheology behaviour of Shanghai soft clay. Hydrogeology & Engineering Geology. 2010; 37: 74-8.

[3] Wu C, Ye G, Zhang L, Bishop D, Wang J. Depositional environment and geotechnical properties of Shanghai clay: a comparison with Ariake and Bangkok clays. Bull Eng Geol Environ 2015; 74:717-32.

[4] Ganesan S, Kuo M, Bolton M. Influences on pipeline interface friction measured in direct shear tests. Geotech Test Journal. 2014;37 37(1):20130008.

[5] Win M, Arulrajah A, Sukmak P, Horpibulsuk S, et al. Mineralogy and geotechnical properties of Singapore marine clay at Changi. Soils and Foundations. 2015; 55(3):600-613.

[6] Zhang XW, Wang CM. Empirical creep model for saturated soft soil. Journal of Central South University (Science and Technology). 2011; 42:791-6.

[7] Song F, Zhao FS. Neural network model for rheology of rock and soil under step loading. Rock and Soil Mechanics. 2006:1187-90.

[8] Chen CF, Liu H, Xiao Y. Ann based creep constitutive model for marine sediment clay. Journal of Engineering Geology. 2008:507-11.

[9] Kong L, Wang YC, Zheng YR. Artificial neural networks and its application in modelling of soil. journal of Ningxia university (Natural Science Edition). 2000:214-6.

[10] Zhou QJ, Chen XP. Research on neural network model of deformation behavior of soft soil. Industrial Construction. 2008:48-53.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230212

Numerical Investigation on Geomechanical Response and Well Performance in Shale Condensate Gas Reservoirs

Wentao LU[a], Yong ZENG[a], Mingming LIU[b, 1] and Lijun LIU[b]
[a] *Research Institute of Exploration and Development, Sinopec Jianghan Oilfield Company, Wuhan 430223, China*
[b] *Energy College, Chengdu University of Technology, Chengdu, 610059, China*

Abstract. Compared with conventional reservoirs, shale gas flow is greatly affected by matrix/fracture deformation, as well as nonlinear coupled transport mechanisms. In this paper, a hydro-mechanical coupled model is presented to describe the fluid flow in deformable shale formation. A unified compositional model is developed for modeling of multiphase fluid flow with phase transition. A series of mechanisms, including Knudsen diffusion, multi-component adsorption, confined phase behavior and molecular diffusion, are considered for accurate description of fluid flow in shale reservoirs. Matrix deformation is based on the linear poroelasticity theory. The fractures with complex geometry are modeled with the embedded discrete fracture model (EDFM). The mechanical responses of fractures are handled by different constitutive models, which are implemented into the coupled model. The flow and geomechanical models are spatially discretized using finite volume method (FVM) and finite element method (FEM), and the sequentially iterative approach is applied for solving the coupled model. Then the impacts of fracture orientation, in-situ stress condition, and bottom hole pressure on the mechanical deformation and gas production are investigated through sensitivity analysis. With multiple mechanisms and dynamic fracture behavior incorporated, the geomechanical response and well performance in shale condensate gas reservoirs can be accurately captured.

Keywords. Coupled flow and geomechanics; FEM; FVM; fracture network; shale condensate gas

1. Introduction

Over the last decade, shale resources have gained great attentions worldwide due to their large reserves. Hydraulic fracturing has been one of the most effective methods to improve hydrocarbon recovery from shale reservoirs [1]. Complex fracture networks can be generated when hydraulic fractures interact with the pre-existing natural micro-fractures [2]. Therefore, the formation can be divided into three components after hydraulic fracturing, which are matrix, micro-fractures, and hydraulic fractures.

[1] Mingming Liu, Energy College, Chengdu University of Technology, Chengdu, 610059, China; E-mail: woxinfeixiang1210@126.com.

Shale rocks are commonly characterized by nanoscale pores and abundant organic matters. These lead to several nonlinear fluid flow mechanisms, such as Knudsen diffusion and adsorption. Many efforts have been made to investigate the effects of these factors on fluid flow in shale reservoirs [3-5]. Among these studies, the hybrid model proposed by Lijun et al.[5] gives a comprehensive description of the multiple mechanisms, which include stress sensitivity, confined phase behavior, adsorption, and Knudsen diffusion. As the major flow conduits, fractures provide conductive pathways inside the formation, and bring a large portion of rock into direct contact with the well [6] . Meanwhile, fractures are highly sensitive to the mechanical loadings. Great aperture and conductivity lose may happen during depletion because of the increase of effective stress on fractures. However, the responses of hydraulic fractures and micro-fractures to the effective stress change are different. A lot of analytical and empirical hydraulic fracture conductivity models have been developed [7-9] Most empirical models depend on the experimental fittings, and the accuracy of predicting the proppant embedment and fracture conductivity may be not satisfactory in some cases. The analytical model derived with the contact mechanics give a unified description and good accuracy, such as Li's model [7]. Different from hydraulic fractures, micro-fractures are usually less propped, whose mechanical behavior mainly depends on its rough surfaces. The Barton-Bandis model [10, 11] is widely used to mimic the natural fracture deformation. Moinfar et al. [12] proposed a coupled flow and geomechanics model for fractured reservoir by introducing the EDFM and empirical joint models, but without the propped hydraulic fracture model. Jiang and Yang [13] developed a model for stress sensitive fractured shale reservoir with the nonlinear transport mechanisms and propped hydraulic fracture constitutive model, but lack of systematic analysis of the effect of fracture dynamic behavior.

In this study, a hydro-mechanical coupled model for shale condensate gas reservoirs is presented. A unified compositional model considering multiple mechanisms is developed for fluid flow in shale reservoirs. A linear elastic constitutive model coupled with propped hydraulic-fracture model and empirical natural-fracture model is used to model the mechanical deformation of matrix and fractures. The complex fracture network is handled by EDFM. Then sensitivity analysis is conducted to investigate the effects of fracture orientation, in-situ stress, and bottom-hole pressure on the shale gas production.

2. Numerical Model

Firstly, the shale condensate gas reservoir is discretized with structured grids. The complex network with hydraulic fractures and micro-fractures are efficiently modeled with EDFM. As shown in Fig 1, the orthogonal structured grids are applied for the matrix region, and fractures grids are generated by segmenting fractures with the matrix grid lines. Then the connectivity among these grids are extracted for the follow-up reservoir simulations.

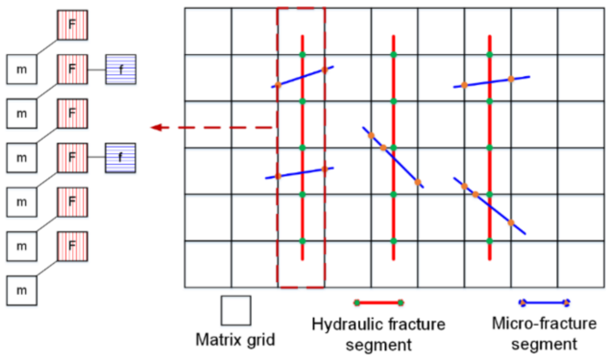


Fig 1. Schematic of shale condensate gas reservoir discretization: (left) grid connection and (right) grid structure.

For flow model, Eq. (1) is numerically discretized through FVM.

$$\begin{aligned} R_n^{i,t+1} = {} & \frac{\{V[\phi(\rho_0 s_o x_i + \rho_g s_g y_i) + q_{ads,i}]\}_n^{t+1} - \{V[\phi(\rho_0 S_o x_i + \rho_g S_g y_i) + q_{ads,i}]\}_n^{t}}{\Delta t} \\ & - \sum_{m\in\eta_n} [(\rho_0 x_i \lambda_o)_{nm+\frac{1}{2}}^{t+1} \gamma_{nm}^{t+1} (\psi_{om}^{t+1} - \psi_{on}^{t+1}) + (\rho_g y_i \lambda_g)_{nm+\frac{1}{2}}^{t+1} \gamma_{nm}^{t+1} (\psi_{gm}^{t+1} - \psi_{gn}^{t+1})] \\ & - \sum_{m\in\eta_n} D_{eff,i} A_{nm}^{t+1} \frac{(\rho_g y_i)_m^{t+1} - (\rho_g y_i)_n^{t+1}}{d_n^{t+1} + d_m^{t+1}} - (V q_i)_n^{t+1} \end{aligned} \tag{1}$$

$$\begin{aligned} R_n^{w,t+1} = {} & \frac{(V\phi\rho_w S_w)_n^{t+1} - (V\phi\rho_w S_w)_n^{t}}{\Delta t} \\ & - \sum\nolimits_{m\in\eta_n} \left[(\rho_w \lambda_w)_{nm+\frac{1}{2}}^{t+1} \gamma_{nm}^{t+1} (\psi_{wm}^{t+1} - \psi_{wn}^{t+1}) \right] (V q_w)_n^{t+1} \end{aligned} \tag{2}$$

where η_n represents grids connected and n is grid block; nm+1/2 is the interface of grid blocks n and m; λ is phase mobility calculated as $\lambda_\beta = k_{r\beta}/\mu_\beta$; γ is the transmissibility calculated as

$$\gamma_{\mathrm{nm}} = \frac{A_{nm} K_{nm+1/2}}{d_n + d_m} \tag{3}$$

where A is the interface area, m^2; d is the vertical distances from cell center to interface, m.

The Newton-Raphson method is applied to solve the nonlinear Eqs. (2)-(3) with the following scheme

$$\sum_p \frac{\partial \mathrm{R}_n^{\beta,t+1}(x_{p,k})}{\partial x_p} \delta x_{p,k+1} = -\mathrm{R}_n^{\beta,t+1}(x_{p,k}) \tag{4}$$

$$x_{p,k+1} = x_{p,k} + \delta x_{p,k+1} \tag{5}$$

where x is the primary variable; p is the variable index; k is the iteration level.

For the geomechanical model, FEM is adopted for the numerical discretization. The weak form of Eq. (2) can be obtained as follows

$$\int_{\Omega}\delta\varepsilon : \sigma d\Omega = \int_{\Omega}\delta u \cdot \rho_b g \, d\Omega + \int_{\Gamma}\delta u \cdot \mathrm{T}_{ett}\, \mathrm{d}\Gamma \tag{6}$$

and Ω denotes the reservoir domain; Γ denotes the computational boundary with fixed traction; $\mathbf{T}_{\mathrm{ext}}$ is the exerted traction on boundary, Pa. The displacement vector $\boldsymbol{u}$ can be calculated with the nodal displacement $\bar{\boldsymbol{u}}$ and interpolation function $\boldsymbol{N}$

$$u = N\bar{u} \tag{7}$$

Substituting Eq. (7) into Eq. (6), the matrix-vector form of Eq. (7) can be obtained as follows

$$K\bar{u} = f \tag{8}$$

where

$$\boldsymbol{K} = \int_{\boldsymbol{\Omega}} \boldsymbol{B}^{\boldsymbol{T}}\boldsymbol{D}\boldsymbol{B}\mathrm{d}\Omega \tag{9}$$

$$\boldsymbol{Q} = \int_{\boldsymbol{\Omega}} \boldsymbol{\alpha p_t B^T m d\Omega} \tag{10}$$

$$\boldsymbol{f} = \int_{\Gamma} \boldsymbol{N}^T\boldsymbol{T}_{ext}d\boldsymbol{\Gamma} + \int_{\boldsymbol{\Omega}} \boldsymbol{N}^T\rho_b g\mathrm{d}\Omega \tag{11}$$

where $\boldsymbol{B} = \boldsymbol{LN}$; $\boldsymbol{L}$ is the matrix consisting with differential operators; $\mathbf{m}$ is delta Dirac function vector. The fixed-stress split method is applied for the solution of the coupled model due to its flexibility and stability[14], [15]. In one time step, the flow model is firstly solved by fixing the total mean stress, and then the updated fluid pressure is transferred to the solution of geomechanical model. The reservoir porosity and permeability update after the geomechanical problem is solved. This procedure repeats in one time step until the fluid pressure and rock deformation becomes stable, then the next time step begins. The flowchart of the solution procedure is shown in Fig 2.

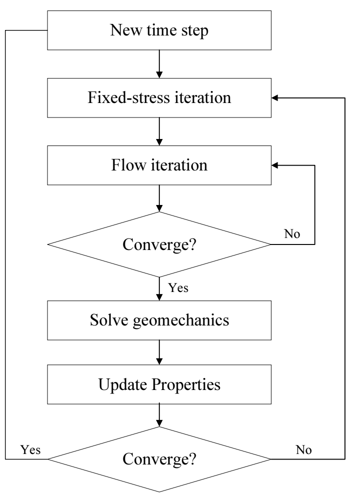


Fig 2. Flowchart of coupled flow and geomechanics problem.

3. Results and Discussions

A multi-stage fractured shale condensate gas reservoir model is used the parameters slightly modified from Roussel et al.[16]. The fluid properties and relative permeability curves in Liu et al.[5] are used. The reservoir and fracture geometry are shown in Fig 3, in which random micro-fracture networks are generated with the density of 0.08m/m^2. Table 1 shows the basic reservoir parameters for the base case.

Based on the reservoir model, the base case is simulated. Fig 4 shows the results of pressure and saturation distribution after 1500 days' depletion. The gas production results with and without consideration of geomechanics are given in Fig 5. It is clear that geomechanics has great influence on the gas production, and without considering geomechanics, the production may be greatly overestimated.

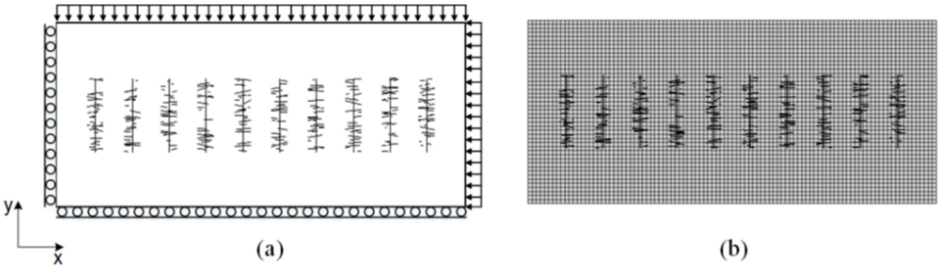


Fig 3. sketch for (a) fractured shale reservoir model and (b) computational grids.

Table 1. Details parameters.

Parameters	Value	Unit
Model dimension	1000×450×10	m
Grid size	10×10×10	m
Matrix porosity	0.1	–
Matrix permeability	6×10^{-4}	mD
Matrix Young's modulus	1.5×10^{4}	MPa
Matrix Poisson's ratio	0.7	–
Initial micro-fracture permeability	5×10^{3}	mD
Initial micro-fracture aperture	1×10^{-3}	m
Initial normal stiffness of micro-fracture	10^{4}	MPa
Maximum closure of micro-fracture	9×10^{-4}	m
Initial hydraulic fracture permeability	7×10^{4}	mD
Initial hydraulic fracture aperture	4×10^{-3}	m
Hydraulic fracture half-length	90	m
Hydraulic fracture spacing	90	m
Proppant Young's modulus	2×10^{4}	MPa
Proppant diameter	3×10^{-4}	m
Initial gas saturation	0.75	–
Initial reservoir pressure	36.03	MPa
Initial reservoir temperature	384.82	K
Well radius	0.1	m
Bottom hole pressure	10	MPa
In-situ stress in x-direction	75	MPa
In-situ stress in y-direction	75	MPa

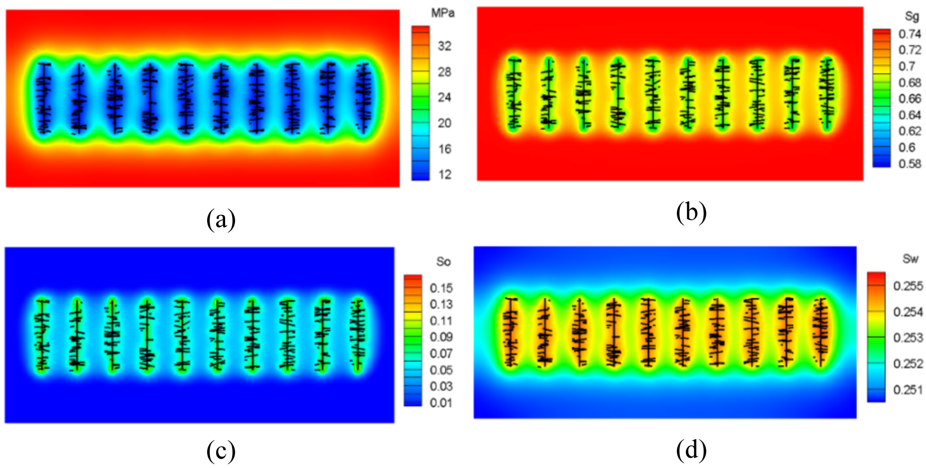


Fig 4. Simulation results of (a) pressure and saturations of (b) gas, (c) oil, and (d) water.

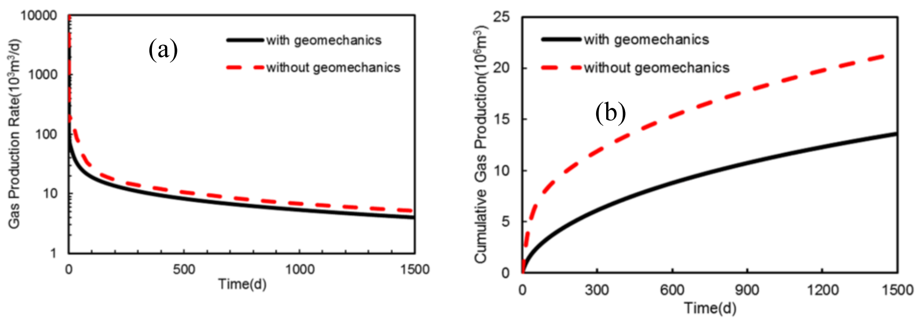


Fig 5. Curves for (a) gas production rate and (b) cumulative gas production with and without geomechanics.

3.1. Effect of Fracture Orientation

Different fracture orientations may occur when the horizontal wells are not drilled along the direction of the minimum principal stress. Therefore, it is necessary to study the dynamic behavior of fractures and gas production with different fracture orientations. With implementation of EDFM, complex fracture geometry can be handled in the simulation. Then different fracture orientations, consisting of 90˚, 60˚, and 30˚ angle of inclination, are investigated. Fig 6 shows the simulation results of pressure distribution. The gas production curves are given in Fig 7.

As shown in Fig 6, all the three cases have different drainage areas, in which the 90˚-inclination gives the largest depleted area, and 30˚-inclination has the least one. The size of depleted area commonly determines the production rate and gas recovery. As shown in Fig 7, as the fracture inclination decreases, the gas production decrease. But there is little difference between the gas production with 90˚-inclination and 60˚-inclination fractures because of the similar size of depleted area. Another observation from pressure distribution is that with the fracture inclination angle decrease, the distance among the fractures also decreases, which leads to the interference of neighboring fracture at late stage and explains why the gas production declines at late stage for the 30˚-inclination fractures.

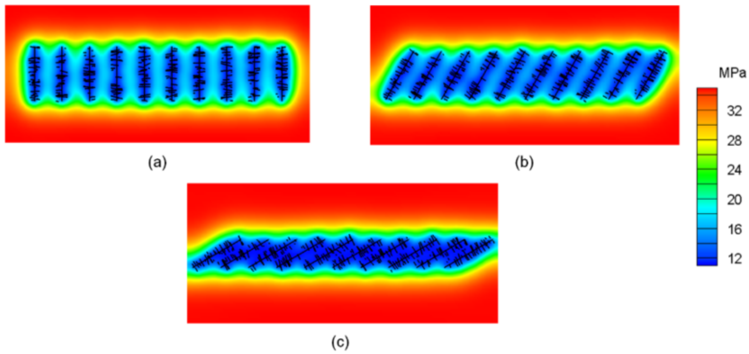


Fig 6 .Pressure distribution (a) 90˚, (b) 60˚, and 30˚.

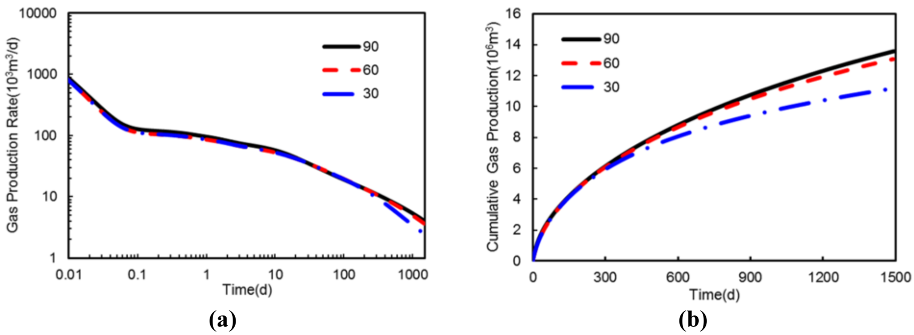


Fig 7. Curves for (a) gas production rate and (b) cumulative gas production with different fracture orientations.

3.2. In-Situ Stress

The effect of in-situ stress, here referring to the two horizontal principal stresses, on the gas production is investigated due to its important role in stress distribution and fracture aperture evolution. The x- and y-direction in-situ stress are analyzed separately, and take the values of 75MPa, which is a base case in the preceding subsections, as well as 55MPa and 35MPa. Fig 8 shows the hydraulic fracture aperture distribution with different x-direction in-situ stress, while the micro-fracture apertures show little difference and are not plotted. Similarly, Fig 9 shows the micro-fracture aperture distribution with different y-direction in-situ stress, and hydraulic fracture apertures are not given. Then the results of gas production are given in Fig 10.

The hydraulic fracture aperture changes, but little, with the x-direction in-situ stress, while the micro-fracture aperture is sensitive to the y-direction in-situ stress change which is shown in Fig 8 and Fig 9. The micro-fracture aperture with 35MPa y-direction in-situ stress is nearly twice of that with 75MPa y-direction in-situ stress. The directional effect of fracture aperture change is related to the normal directions of fracture surface. Obviously, in this case, the hydraulic fractures and micro-fractures are sensitive to the x- and y-direction in-situ stress, respectively. As for the different degree of change for hydraulic fracture and micro-fracture aperture, it is related with their different mechanical properties. The hydraulic fractures are stiffer due to the good propping of proppants rather than the weak rough surfaces. As a result, the gas production, shown in Fig 10, increases with the y-direction in-situ stress decreasing, while changes little with the y-direction in-situ stress. These results may guide the well drilling and completion that it is important to keep the micro-fracture networks suffering less stress.

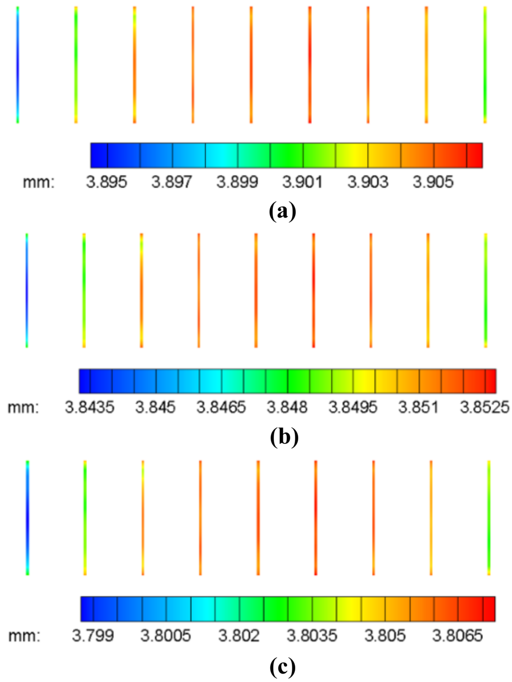


Fig 8 . Hydraulic fracture aperture with different x-direction in-situ stress of (a) 35, (b) 55, and 75MPa.

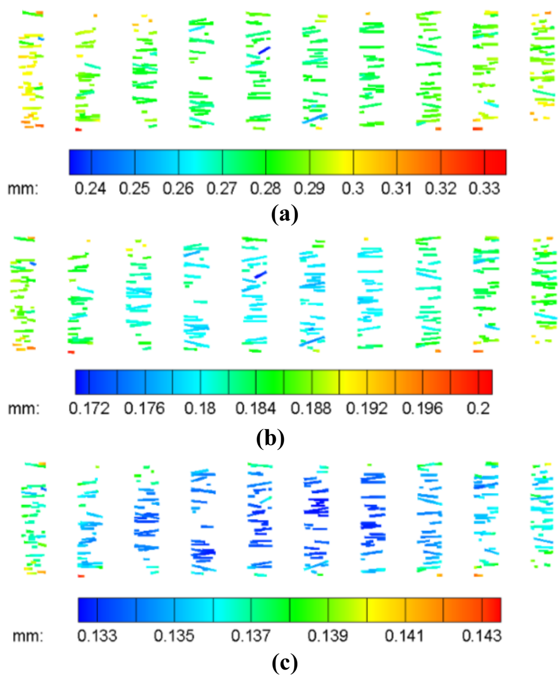


Fig 9. Micro-fracture aperture with different y-direction in-situ stress of (a) 35, (b) 55, and 75MPa.

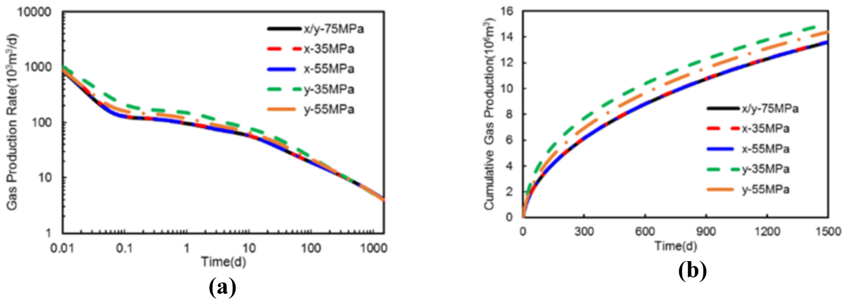


Fig 10. Curves for (a) gas production rate and (b) cumulative gas production with different in-situ stress.

4. Conclusions

A hydro-mechanical coupled model to simulate the geomechanical response and well performance in shale condensate gas reservoirs was developed in this work. A unified compositional model considering multiple mechanisms is implemented. The EDFM is adopted for efficient modeling of complex fracture geometry, and the mechanical response of micro-fractures and hydraulic fractures is handled by different constitutive models. Then the effects of fracture orientation, in-situ stress, and bottom-hole pressure on the gas production are investigated. The following conclusions drawing from the results are obtained:

1. Significant fracture aperture reduction and matrix deformation are obtained under geomechanical effect, and gas production can be greatly overestimated without considering this effect.
2. With the fracture inclination angle decreasing, the interference happens among fractures, leading to reduction in gas production.
3. The gas production is sensitive to the in-situ stress in the direction normal to micro-fracture because micro-fractures are more stress sensitive.
4. The gas production decreases greatly with the increase of bottom-hole pressure.

References

[1] Hubbert MK, Willis DG. Mechanics of hydraulic fracturing. Transactions of the AIME, 1957, 210(01): 153-168.
[2] Ahn CH, Chang OC, Dilmore R, Wang JY. A Hydraulic Fracture Network Propagation Model in Shale Gas Reservoirs: Parametric Studies to Enhance the Effectiveness of Stimulation. SPE/AAPG/SEG Unconventional Resources Technology Conference. OnePetro, 2014.
[3] Moridis GJ, Blasingame TA. and Freeman CM. Analysis of Mechanisms of Flow in Fractured Tight-Gas and Shale-Gas Reservoirs. SPE Latin American and Caribbean Petroleum Engineering Conference. 2010, Lima, Peru, Society of Petroleum Engineers: 22.
[4] Wu YS, Li J, Ding DY, Di Y. A generalized framework model for the simulation of gas production in unconventional gas reservoirs. Spe Journal, 2014, 19(05): 845-857.
[5] Lijun L, Jun Y, Hai S, Zhaoqin H, Xia Y, Longlong L. Compositional modeling of shale condensate gas flow with multiple transport mechanisms. Journal of Petroleum Science and Engineering, 2019, 172: 1186-1201.
[6] Shakiba M, Sepehrnoori K. Using embedded discrete fracture model (EDFM) and microseismic monitoring data to characterize the complex hydraulic fracture networks. SPE annual technical conference and exhibition. OnePetro, 2015.

[7] Li H, Wang K, Xie J, Li Y, Zhu S. A new mathematical model to calculate sand-packed fracture conductivity. Journal of Natural Gas Science and Engineering, 2016, 35: 567-582.
[8] Li K, Gao Y, Lyu Y,Wang M. New mathematical models for calculating proppant embedment and fracture conductivity. SPE Journal, 2015, 20(03): 496-507.
[9] Chen D, Ye Z, Pan Z, Zhou Y, Zhang J. A permeability model for the hydraulic fracture filled with proppant packs under combined effect of compaction and embedment. Journal of Petroleum Science and Engineering, 2017, 149: 428-435.
[10] Bandis S C, Lumsden A C, Barton N R. Fundamentals of rock joint deformation. International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts. Pergamon, 1983, 20(6): 249-268.
[11] Barton N, Bandis S, Bakhtar K. Strength, deformation and conductivity coupling of rock joints. International journal of rock mechanics and mining sciences & geomechanics abstracts. Pergamon, 1985, 22(3): 121-140.
[12] Moinfar A, Sepehrnoori K, Johns RT, Varavei A. Coupled geomechanics and flow simulation for an embedded discrete fracture model. SPE reservoir simulation symposium. OnePetro, 2013.
[13] Jiang J, Yang J. Coupled fluid flow and geomechanics modeling of stress-sensitive production behavior in fractured shale gas reservoirs. International Journal of Rock Mechanics and Mining Sciences, 2018, 101: 1-12.
[14] Kim J, Tchelepi HA, Juanes R. Stability and convergence of sequential methods for coupled flow and geomechanics: Fixed-stress and fixed-strain splits. Computer Methods in Applied Mechanics and Engineering, 2011, 200(13-16): 1591-1606.
[15] Dana S, Ganis B, Wheeler MF. A multiscale fixed stress split iterative scheme for coupled flow and poromechanics in deep subsurface reservoirs. Journal of Computational Physics, 2018, 352: 1-22.
[16] Roussel N P, Florez H A, Rodriguez A A. Hydraulic fracture propagation from infill horizontal wells. SPE annual technical conference and exhibition. OnePetro, 2013.

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230213

Spoken Word List Development for Children's Second Language Study Based on Expression Category

Jingyi XIE[1]
ICCS, Fujian Normal University, Fuzhou, China, 350117

Abstract. The study of spoken word lists is critical for the Chinese language learning by foreign children. At the same time, spoken corpora of native children's language provide an important basis for developing such word lists. Word list development is based on more than word frequency: other important factors must be considered. By analyzing a corpus of children's spoken language, this study proves that the aggregation of children's spoken vocabulary is related to expression categories, as well as to their native language. Furthermore, children's spoken vocabulary changes with age as follows: generalization gradually disappears; the number of words increases, and their meaning changes; a functional change in the use of sentences with common words occurs. The paper concludes by summarizing some core recommendations for children's spoken word lists: gathering words according to themes; dividing polysemy and collocation blocks according to age; adding and subtracting words according to different native language backgrounds.

Keywords. Spoken word list, children's language development, children's Chinese proficiency grading standards.

1. Introduction

1.1. Research Background

There has been an increase in young Chinese learners overseas in recent years. The average percentage of young learners is over 50%, with some countries reaching or exceeding 60% [1]. However, children's cognition and use of language are distinct from those of adults, making children's learning process of Chinese as a second language different and unsuitable to the standards applied to adults. Alternatively, the approach to teaching children a second language should be based on the same language expression system found among native speakers of the same age.

Research on children's brain and language development suggests that their first language begins to mature at three, and they establish a basic grammar system by age 5 [2]. The language and range of expression produced by children speaking their mother tongue at this age can be a reference for children learning that tongue as a second language [3].

[1] Corresponding Author: Jingyi Xie, International College of Chinese Study, Fujian Normal University, Fuzhou, China, 350117. Email: xjyfjnu@fjnu.edu.cn

This article addresses the corpora of children's spoken language to uncover the word distribution and development patterns. It discusses methods for developing children's second-language spoken word lists from the following perspectives: (1) the lexical distribution characteristics of children's spoken language; (2) the relationship between children's second and native languages; and (3) the development characteristics of children's languages.

1.2. Related Research

In this section, previous research findings regarding the development of children's spoken word lists are discussed in terms of the relationship between word lists and word frequency, lexical classification and the clustering of children's spoken words, and the development of meaning in children's spoken language.

The relationship between word list development and word frequency. Several language experiments have confirmed the effect of word frequency in the target language on second language learning [4]. There was general agreement that high-frequency words in the target language should be mastered first [5]. Dolch's "A Basic Sight Vocabulary", a formative text in the international community, was also based on word frequency. However, word lists based purely on word frequency have certain defects. Research regarding Chinese second-language teaching has also suggested that frequency statistics should not be limited to word frequency but should rather be expanded to encompass meaning [6].

The clustering of words in categories within children's spoken vocabulary. The linguistic expressions used by children aged 3 to 5 are closely related to the needs of daily life, and the vocabulary used to express these needs can be aggregated to form spoken language expression categories. Many children's word lists are also organized according to the clustering of their spoken vocabulary [7]. In best-selling books designed for children to learn English vocabulary, the vocabulary is divided into 28 categories, such as daily necessities, food, furniture, toys, clothing, kinship figures, places, natural phenomena, substances and materials, weather, seasons, plants, animals, time and date, body parts, orientation nouns, properties of things (colors and shapes), proper nouns, occupations, sports, thinking and abilities, transportation, and holidays and traditions. This is the basis for categorizing the vocabulary later on, as shown in studies [8-10].

The development of meaning systems in children's oral expressions.

Children's language development is intimately related to their cognitive development. Their semantic systems and the words they represent develop with age. Concerning children's acquisition of polysemic verbs, a positive correlation has been found between the time of acquisition and the frequency of use [11]. They will likewise use high-frequency tenses (generally prototypes) to refer to low-frequency tenses directly. For example, the acquisition of children's modal verbs shows a strong developmental trend that works from "root modality", especially "dynamic modality", to "cognitive modality" [12].

2. Study Background and Approach

The paper draws upon various corpora containing videos of daily spoken interactions between Chinese children and adults. After transcription, the combined corpus was divided into turns, and the number of words was counted. The results incorporate more than 20,000 turns and

280,000 words. The primary sources for the combined materials were the transcribed audio and video-recording-based corpus CHILDES (Child Language Data Exchange System).[2] The recordings of children whose native language was Chinese were selected. The study covered 14 children, aged 4.3 to 5.9 years old, from Beijing, Nanjing, Hangzhou, and other places in China. The corpus mainly contained conversations between children, parents, and researchers undertaking activities such as playing games, reading picture books, and telling stories. We used a total of about 30 hours of transcribed recordings. After the segmentation of the recordings, 16,960 turns were obtained, with a total of 121,609 words.

2.1. The clustering of high-frequency words in different categories of children's oral expressions

In the oral expressions children use in everyday life, there is often a connection of meaning and topic between high-frequency verbs and high-frequency nouns. An overall analysis of the corpus also showed that high-frequency words covered a wider range of the content than low-frequency words but that the trends relating to the coverage of the verbs and nouns differed. The coverage of high-frequency nouns grew slower than that of verbs. Concerning the general and word list-based corpora, the top 6.93% and 32.31% of most frequent nouns covered only 50% and 80% of the expressions, respectively. The nouns covered a total of 28 categories, among which the high-frequency nouns were clustered in categories relating to daily necessities, food, places, and natural phenomena, which are closely related to children's lives. These exceeded 50% of all nouns in the main corpus and 80% of all nouns in the word list-based corpus, so these comprise the main components of noun vocabulary in children's oral expressions.

2.2. Commonalities and differences in how vocabulary is used in different categories of children's oral expressions

There are core elements of human language that are universal. Word clusters that express basic conceptual categories, such as people, food, shapes, weather, etc., exist in almost all languages. Children all acquire language in the context of particular social circumstances. During this process, they are inevitably affected by their specific linguistic and socio-cultural environment, which will differ from place to place.

According to our analysis, there were 22 verbs representing actions and behaviors in the English children's spoken corpus. In comparison, there were just 14 verbs in the oral expressions of Chinese native-speaking children. Nonetheless, some notable overlaps existed between the Chinese and English high-frequency verbs. Another point worth noting is that food nouns such as "rice" (*mifan*), "noodles" (*miantiao*), and "soup" (*tang*) in Chinese are not present in Dolch's "Noun List".

[2] https://childes.talkbank.org/

3. Longitudinal Analysis of the Development of Word Meaning Systems in Children

3.1. Expansion in the number of high-frequency verbs

As the number of words in children's oral expressions increases with age, the number of words in children's oral word lists must also expand as they age. However, this expansion should involve increased quantity, concretization, and refinement of vocabulary content.

The performed analysis of the corpus revealed that the range of expressions covering different functions in children's spoken language reached 85% of that of adults by the time they were 5. Children's vocabulary growth was not just learning 31 more words: their expression categories exhibited a permanrent dynamic change. For example, "want" (*yao*) in the children's spoken language corpus could be matched with almost any noun, e.g., "baby wants red"; "little mouse wants mother"; "I want rice"; and so on. These expressions in adult spoken language correspond to such forms as "give me the red one", "little mouse misses its mother", and "I need a bowl of rice", respectively. Therefore, it can be seen that thwe verb "want" (*yao*), in younger children's oral expressions, "shares" meaning and functionality with such verbs as "give" (*gei*), "miss" (*xiang*), and "need" (*xuyao*).

3.2. Changes in the function of polysemic verbs in children's oral expressions

When children and adults use the same verbs, there are differences in the distribution of functions being expressed. The choice of a certain category of expression and its representative words needs to be modulated according to age [13]. For example, at the age of 3 years, there is no need to pay special attention to selecting categories related to guiding others to complete a message. This choice becomes more relevant for children aged five and above. At the same time, when guiding others, adult educators should be aware that their common vocabulary differs from children's. For example, adults like to give direct orders to others, e.g., "Go eat..." and "Go get...", while children may prefer to say "I want...", which is equal to "Help me to...".

3.3. Vocabulary growth and changes in the meaning categories of nouns

As a child's overall vocabulary increases, the range of words in different sense categories sometimes grows at different rates. We found four patterns in the growth of noun vocabulary. The proportion of nouns in the "daily necessities" category steadily declined (from 26.6 to 11.9%). The proportion of nouns in the category of "natural phenomena" skyrocketed dropped from 2.9 to 18.6% [3]. "Time and date" nouns increased rapidly between 3 and 4 years of age but remained more or less the same between 4 and 5 years (the usage frequencies were 2.9, 9.1, and 9.0%, respectively). The proportion of nouns in the "relatives and friends" category remained almost unchanged (the frequencies were 9.8, 9.4, and 9.3%, respectively). Thus, when screening high-frequency words for their use in word lists, it is important to consider what changes occur in children's attention span and cognition as they get older, the implications for changes in categories, and how they select corresponding words.

[3] The increase in the frequency of "celestial nature" nouns here is partly because the words "heaven", "earth", and "star" are polysemous in Chinese, and children acquire more of their meanings as they grow up, leading to an increase in their usage frequency.

4. Principles for the Compilation of CSL Word Lists for L2 Children

The analysis of children's spoken language corpora in this paper shows that in developing CSL Word Lists for children, such parameters as age and cultural background need to be considered in addition to word frequency.

Based on the above findings, we propose the following recommendations regarding the design of word lists:

4.1. The development of word lists should be based on the age characteristics of children

Children's language has age-dependent characteristics, so child-oriented teaching, exam syllabuses, and word lists must be matched with their level of language development. For example, we have found some differences in the frequency distribution of verbs between adults and children. This suggests that when compiling a relevant syllabus for teaching children Chinese verbs, their level of language development must be considered, in addition to the survey results related to global word frequencies.

Concerning the syllabus and teaching content organization, it is impossible to arrange the teaching sequence according to word frequency directly. Most Chinese verbs have a polysemic structure, e.g., the auxiliary verbs and the structures associated with the verb "is" (*shi*). In teaching these verbs and related structures, their topic and function must be combined, and attention must be paid to the recurrence of high-frequency words with different meanings. A more appropriate teaching sequence needs to be developed that properly recognizes the functional preferences in children's spoken expressions.

4.2. The development of word lists should meet the needs of individualization and localization

This research confirms that children's common vocabulary revolves around their everyday sphere and has a distinctly individual character. Children only care about their own information at this stage of cognitive development. Therefore, to arouse their interest, starting with things around them is necessary. One can then move on to promoting cognitive development, improving cognitive ability, and ultimately improving language ability [14]. Therefore, when planning the instruction content for children and developing a list of bilingual words, one should consider the characteristics of their daily life. The geographical area where the children are located, the focus of their family life, the content of their daily play, etc., can be used as objects in the planning of their teaching. A reasonable next step should be developing various teaching strategies around "can-do" lists to form local vocabulary lists. Beyond this, in the actual use of vocabulary in various countries, the specific content must also be adjusted to reflect the customs and habits of the country where the children are located.

5. Conclusion

Given the findings of this paper, it is not advisable to use adult corpora to determine children's second language learning word lists. The performed analysis of children's native language corpora, the clustering of children's use of verbs and nouns, and differences in children's word choice in different native language contexts revealed significant differences between children's and adults' vocabulary use. The follow-up study envisages expanding the analysis of children's vocabulary and further exploring the development of comprehension, the expression of different functions, and the use of grammatical categories to compile more scientifically grounded word lists that would allow children to accomplish richer functional expressions with fewer words.

References

[1] Li Y.M. Reflections on the younger age of overseas Chinese learners. Chinese Teaching in the World (2018).
[2] Liu L. Q., Zhang H. L. W., Effects of bilingual environments on children's early language development, Chinese Language Strategies, December 2022.
[3] Bergen D., Coscia J. Brain Research and Childhood Education: Implications for Educators, Parents, and Society (Chinese Edition). Assn for Childhood Education Intl, 2001, ISBN-13: 978-0871731548
[4] Li Y.M. The development of children's language. Wuhan: Huazhong Normal University Press, 1995.
[5] Farrell L., Osenga T., Hunter M., Comparing the Dolch and Fry High-Frequency Word Lists, Readstters,
[6] Li Y. M. Common destiny, common language. "Guangming Daily", 2020/08/22.
[7] Zhang Y.Q., Zhou J. S., Fu J. Acquisition strategies of polysemy in early Chinese children: A case study of Beijing dialect children. Chinese Language and Literature (2010).
[8] Zhang Y. Q., Liang Y.X. The early development of children's modal semantics and the significance of demonstration. Chinese Learning (2021).
[9] Ellis N.C., Simpson-Vlach R., Raynard C., Formulaic language in native and second language speakers. Psycholinguistics, corpus linguistics, and TESOL (2008).
[10] Krashen. S. and Terrell, T.D. The Natural Approach. Prentice Hall Europe (1998).
[11] Widen S. C. & Russell J. A. A closer look at preschoolers' freely produced labels for facial expressions. Developmental Psychology (2003).
[12] Cui Na. Discussion on classification and grading of polysemous morphemes in teaching Chinese as a foreign language. Education and Teaching Forum (2019).
[13] Michael A. K. Halliday. Infant language(Chinese Edition). Peking University Press, Beijing (2015).
[14] Yan Y.M. Introduction to Information Science. Wuhan University Press, Wuhan (1994).

Modern Management Based on Big Data IV
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA230214

Exploring the National First-Class Specialty Construction of Pharmaceutical Engineering Under the Background of "New Engineering" Project

Zhenhua CHEN[a], Zekai CHEN[a], Liangliang ZHOU[a], Pengwu ZHENG[a], Yuyu XU[1,b]
[a] *School of Pharmacy, Jiangxi Science and Technology Normal University, Nanchang 330013, China*
[b] *College of Tourism and History Culture, Jiangxi Science and Technology Normal University, Nanchang 330022, China*

Abstract. The present study envisages the construction and reforming practice of the national first-class specialty in local colleges under the background of the "New Engineering" project. Specifically, the construction of engineering teaching team, and the construction of practical teaching platform within and out of universities have been explored. Furthermore, the revolution of the teaching method and medium and the cultivation of innovative talents has been discussed. With the implementation of the "New Engineering" reform and other professional construction initiatives, the quality of pharmaceutical engineering education in the School of Pharmacy, Jiangxi Science and Technology Normal University has significantly improved, and the influence and reputation of the specialties has been significantly enhanced. As a result, the Pharmaceutical Engineering major of the Jiangxi Science and Technology Normal University has been rated as a first-class advantage specialty in Jiangxi Province in 2022, and has ranked first in the comprehensive evaluation of majors in Jiangxi Province for two consecutive rounds.

Keywords. New engineering; Pharmaceutical engineering; First-class specialty

1. Introduction

In the present global scenario, a new round of scientific and technical revolution accompanied with the industrial transformation has accelerated, thereby leading to an intensified competition in the comprehensive national strength among different countries. To accelerate the construction of the new engineering specialty, and facilitate the economic transformation and upgradation, there is an implementation of a series of ongoing significant strategies such as "Innovation-driven Development", "One Belt, One Road", "Internet Plus", and "Made in China 2025". To pave a way in the vigorous development of the new economy characterized by new technology, there is a need of innovative forms of business and industries. In this pursuit, there is an inevitable need

[1] Corresponding Author: Yuyu Xu, College of Tourism and History Culture, Jiangxi Science and Technology Normal University, Nanchang 330022, China; E-mail: 55762031@qq.com.

of engineering talents with a higher ability of innovation, entrepreneurship and cross-border integration.

In order to satisfy the strategic demand of new engineering talents in the new economy and new industries, the Ministry of Education of China has put forward the "New Engineering" action plan of higher education of engineering in February 2017. This plan aims to cope with changes and shaping the future using an approach of inheritance and innovation, cross and fusion, and coordination and sharing. This can cultivate diversified and innovative outstanding engineering talents for providing intellectual security and talent support in the development of the new economy and new industry [1].

According to the basic characteristics of the new industry and the basic requirements of the new engineering, the main shortcomings of the existing talent cultivation system in most local engineering colleges are the insufficient openness and integration towards industries, and the weakness of adaptability and support. These can be specifically described as: (1) The developments in the positioning and connotation of universities are not compatible with the national economic development strategies and industrial development needs; (2) The talent cultivation system does not adapt to the industrial development; (3) The school-enterprise interaction system cannot meet the needs of the talent cultivation and collaborative innovation in the context of the new industry; (4) Implementation of discipline and scientific research activities do not satisfy the industrial demands and feedback to talent cultivation; (5) The training system cannot effectively support the achievement of the goal of the new engineering talent cultivation; (6) The university governance system cannot meet the needs during the construction of the new engineering system.

Considering the local undergraduate institutions of the education level, and its positioning of training applied talents for the local area, the present study thoroughly leveraged the advantages of the numerous high-quality practical base and rich experience in school-enterprise cooperation, under the basis of the preliminary pharmaceutical category cultivation [2]. This can be achieved by a proper combination of the actual situation of the major and with the guidance of the demand of modern pharmaceutical companies [3]. Further, by employing the innovative talent training modes, such as the integration of production-engineering education, cooperation between the school-enterprise, promoting the reform of curriculum system and teaching evaluation, and exploring the construction and reform of the national first-class specialty in local colleges, the aforesaid goals can be realized [4-6].

2. Construction of the Teaching Group

A teaching team with a strong practical experience of engineering is a significant condition for ensuring the requirements of the "New Engineering" implementation. Although our teaching group has a combination of the subject expertise and strong engineering practical abilities, the School of Pharmacy has employed various measures to build an engineering teaching team that comprehends the development of the pharmaceutical industry and satisfies the needs of engineering talent cultivation, in order to successfully implement the "New Engineering" project and to construct the national first-class specialty. The specific measures are shown in the Figure 1.

Figure 1. Measures adopted by the engineering teaching team.

3. Construction of the Engineering Training Center

3.1. Construction of Engineering Training Center conforming to GMP standards

In order to enable the students to see and use the common manufacturing facility of the pharmaceutical companies and familiarize its working principle, our institute has built a GMP-compliant oral solid preparation engineering training center, equipped with advanced equipment containing automatic control system and air purification system. This training center is able to simulate the actual production scenarios of the pharmaceutical companies, so that students can master the core engineering literacy and practical aspects of pharmaceutical engineering. Furthermore, this training not only fulfills the internship and practical training needs of students, but also undertakes the extracurricular scientific and technological innovation activities and various scientific and technological competitions.

3.2. Construction of Virtual Simulation Training Center

Construction of virtual simulation training center is an essential development in the virtual simulation teaching, which enables the students to grasp a better understanding of the real production environment, process, and principles in real pharmaceutical industries, through a highly immersive teaching mode. With the support of the Ministry of Education’s collaborative practice project for industry-education integration, our institute has constructed a virtual simulation training center that covers the virtual experiments on solid oral dosage forms, injections, chemical drugs, and traditional Chinese medicine production, through independent research and cooperation with enterprises. Together with the engineering training center and experimental teaching center, it has formed a combined virtual and real on-campus practical teaching platform for pharmaceutical engineering.

4. Construction of Practice Teaching System outside school

4.1. Establishing the Assurance management agency employing the "government-school-industry-enterprise" cooperation

The institutes should establish a structurally sound, clearly managed and assigned authorized representative cooperative organization with the government departments and industry enterprises. This can be an important factor in ensuring the implementation and development of the bidirectional integration of production, and can guarantee the establishment of stable off-campus internship bases. In 2009, the School of Pharmacy, Jiangxi Science and Technology Normal University established a professional teaching guidance committee with experts from the industry and enterprise as the main members. For achieving a better work-performance, experts from pharmaceutical companies in other provinces were invited to participate in the committee and revise the committee charter to clarify the responsibilities and rights of the committee. The committee was responsible for studying the requirements and trends of industry and professional development, and determining the objectives of the talent training. Other objectives of the committee included, determining the knowledge and ability structure of graduates, coordinating the management of on-campus and off-campus internships and job training, and carrying out scientific research and training, and recommending graduates.

4.2. Establishing the Evaluation Standards for Off-Campus Internship

It is essential to regularly launch the evaluation and dynamic adjustment of off-campus internship bases to ensure the quality of internships. Feedback should be provided to the internship bases that fail to meet the evaluation standards, and their qualifications may be canceled if a serious problem appears. The evaluation process should be standardized that has practical and feasible evaluation standards, generally including the following contents: (1) Essential facilities: The internship base should be equipped with the teaching, experimental, internship, and scientific research facilities and other hardware conditions necessary for the student internships. (2) Organized Team: The internship base should lay strict rules and regulations to ensure that students' learning in the internship base is productive. (3) Appropriate Guidance: The internship management and guidance personnel receiving a higher education have a certain amount of theoretical and professional knowledge, and have a rich practical experience and professional skills. Such personnel should be arranged by the internship base to charge and guide the students to study. (4) Undertaking Assessments: The internship base should summarize the internship situation of the interns and evaluate their comprehensive performance during the internship, which provides a basis for the school to evaluate the students' internship performance.

4.3. Establishing a Collaborative Education Model with the Combination of "Industry-University-Research"

By establishing the school-enterprise cooperative internship bases, enterprises provide internship conditions to the schools and create a good internship environment for students. Furthermore, the colleges actively combine the "Industry-University-

Research" cultivation model and cooperate with the off-campus internship bases in accordance with the principle of complementary advantages and mutual benefit. This transforms the scientific and technological research and the talent advantages into productivity, thereby improving the overall competitiveness of the enterprises. Meanwhile, the practice and financial advantages of the off-campus practice base should be utilized to transform it into a test base for product development and scientific research promotion of colleges and universities. The school-enterprise cooperation training can sustainably develop only through the complimentary interaction between the schools and enterprises.

5. Methods and Measures of Improving Teaching

The implementation of the "New Engineering" reform is a very systematic project, in which the reform of the teaching methods and means is a vital factor. The measures to carry out the reforms by the project team are summarized in Figure 2.

Figure 2. Reform measures of teaching methods and means.

5.1. Heuristic and Discussion Teaching Model

By implementing the heuristic and discussion-based teaching methods, the initiative and enthusiasm of the students can be mobilized for learning, and an ability to ask and solve problems can be cultivated. In engineering courses, the assignments mainly consist of small test papers that stimulate the students' active thinking, and cultivate their ability to comprehensively apply knowledge and self-study skills.

5.2. Enterprise Production Instance Teaching Model

Based on their practical learning experience from the industry, the teachers can introduce new technologies used in industrial production process into the curriculum, highlighting the practicality of the content. As a result, the students can understand the principles and concepts in real examples. For example, the Pharmaceutical Technology course introduces the process of biological, chemical, and traditional Chinese medicine production, wherein typical examples come from the real production scenarios in pharmaceutical enterprises.

5.3. "Going out and Inviting in" Teaching Model

"Inviting in" means hiring part-time teachers from the pharmaceutical enterprises with strong practical engineering aspects as the main lecturers for professional courses, who can be responsible for teaching the contents that are closely related to production reality. Meanwhile, the theoretical part of the content can be taught by the school's professional teachers. In each semester, production technology experts from well-known pharmaceutical or cross-national pharmaceutical enterprise can be invited to the university to give special lectures on new pharmaceutical technologies for students.

"Going out" implies the content closely related to the pharmaceutical enterprise production scenes for professional courses that is arranged in the pharmaceutical enterprises with on-site learning. Herein, the enterprise engineering and technical personnel are invited to give on-site lectures to the students.

5.4. Campus Virtual Simulation Experiment Teaching Model

By utilizing the school's engineering training center which meets GMP standards and launching virtual simulation experiment, students should be allowed to personally operate and control the production equipment and become familiar with its working principles. This resolves the problem of not being able to operate equipment due to the mere industrial-visit of the pharmaceutical enterprises. After such training and simulation, the internship in enterprises can achieve the effect of twice the results with half the effort.

5.5. Campus Engineering Practice and Training Teaching Model

Practical teaching and engineering training can be launched by relying on the school's built-in engineering training center as an important training platform, which can allow the students to carry out engineering practice operations before going to the factory for internship. This helps in providing an essential guarantee for improving the students' engineering practical and innovative abilities.

6. Cultivation of Innovation Awareness and Ability

Cultivating innovation awareness and ability is an important aspect of the "New Engineering" reform. While strengthening the cultivation of practical engineering aspects, the cultivation of students' engineering innovation ability can be strengthened from multiple aspects, including laboratory openness, undergraduate innovation, entrepreneurship training, and participation in various academic competitions (such as the international "Internet Plus" university student innovation and entrepreneurship competition, the National College Student Pharmaceutical Engineering Design Competition). Meanwhile, regular invitations can be given to alumni and expert scholars who have achieved/provided outstanding contribution to scientific research to deliver academic lectures on scientific and technological innovation. Also, the technical engineers from pharmaceutical companies and pharmaceutical design institutes can be invited to give special reports on engineering innovation. The specific measures that can be taken are shown in Figure 3.

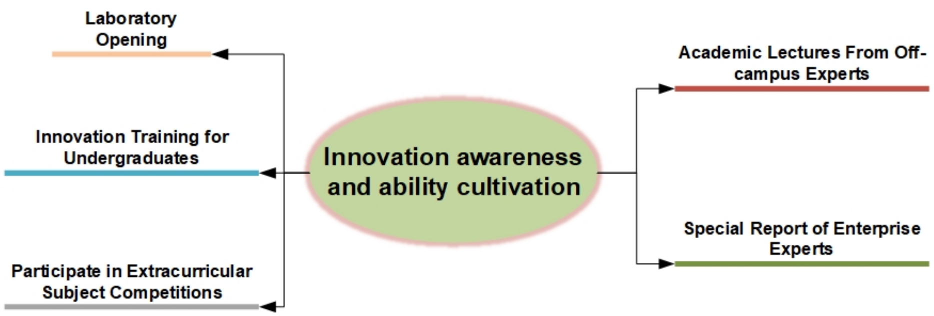


Figure 3. Measures of cultivating the innovative ability of college students.

7. Conclusion

With the implementation of the "New Engineering" reform and other professional construction initiatives, the quality of pharmaceutical engineering education in the School of Pharmacy, Jiangxi Science and Technology Normal University has significantly improved, and the influence and reputation of the specialties has been significantly enhanced. Herein, the students' engineering practical ability has been significantly enhanced and the overall satisfaction of the specialty has kept in high level. There is a shortage of graduates concerning the employment, and the selected candidates have been generally highly evaluated from the employers. Meanwhile, the level of professional construction has been significantly improved. The Pharmaceutical Engineering major of the Jiangxi Science and Technology Normal University has been rated as a first-class advantage specialty in Jiangxi Province in 2022, and has ranked first in the comprehensive evaluation of majors in Jiangxi Province for two consecutive rounds. Further, it has been approved as a five-star major in the Jiangxi Province.

Without doubt, the implementation of "New Engineering" is unmanageable without the participation of teachers. Teachers need to devote a lot of energy and time to education, and education-reform related work. It is a prerequisite for ensuring the construction of first-class specialty and the quality of talent cultivation needed by the government management departments and schools. Relevant policies need to be introduced to encourage the teachers to voluntarily invest more energy into education and teaching.

8. Acknowledgement

This study was supported by the “New Engineering” Research and Practice Project of the Ministry of Education (Letter from the Department of Education and the Higher Education Bureau (2020) No.23) and Jiangxi Provincial Teaching Reform Research Project (JXJG-21-10-22).

References

[1] Lin J. The construction of China’s new engineering disciplines for the future. Tsinghua Journal of Education, 2017 Mar; 38(2): 26-35.
[2] Yin DM, Shan YX. Practical research on the construction path of pharmaceutical engineering characteristic specialty in local colleges under the background of new engineering. Journal of Higher Education, 2021 Oct; 7(30): 36-41.
[3] Liang H, Zhou Y, Qin PY, Zheng GJ, Yuan QP. Building a New Mode of School-enterprise Cooperation Talents Training Based on Emerging. Higher Education in Chemical Engineering, 2021 Aug; 38(04): 120-124.
[4] Bai J, Zhang GS, Shi LL, Tao AL, Dai Y. Research on the Education and Training Mode of Outstanding Pharmaceutical Engineers in Application-Oriented Private Colleges Based on Emerging Engineering. Anhui Chemical Industry, 2022 Aug; 48(04): 130-132.
[5] Ye ZW, Li JZ, Chen Q. Research on the Training Mode of Applied pharmaceutical Engineering Talents under the Background of New Engineering. Pharmaceutical Education, 2022 Jun; 38(03): 19-22.
[6] Ni XC, Yang X, Lu Y, Kong WJ, Qu F, Meng R, Xu LJ, Sun HB, Zhao WH. Exploration on the Training Mode of Innovative Ability of Applied Engineering Professionals under the New Engineering Concept. The Theory and Practice of Innovation and Entrepreneurship, 2022 Nov; 5(23): 130-132.

Subject Index

This page intentionally left blank

Author Index