

# **International Journal of Vocational and Technical Education**

SEPTEMBER 2021 VOLUME 1 NUMBER 3

Publisher: ACADEMIC PUBLISHING HOUSE  
Address: Quastisky Building, Road Town, Tortola, British Virgin Islands  
UK Postal Code: VG1110

E-mail: editor01@acadpubl.com  
<http://www.acadpubl.com/ijvte.aspx>





# CONTENTS

<b>A Survey of Human Pose Estimation Based on Lightweight Neural Network .....</b>	<b>1</b>
<i>Kuo Fu, Xia Liu, Qiyan Yan</i>	
<b>The Trinity of Government School and Enterprise Advantages and Elements Sharing Collaborative Innovation Center .....</b>	<b>4</b>
<i>Ding Wenhui</i>	
<b>Study on the Influences of Application Colleges and Universities Teachers on Ten Education System Construction .....</b>	<b>8</b>
<i>Zhang Yijun, Yuan Dan</i>	
<b>Talking about the Application of Tea Culture in the Management of Class Moral Education in Vocational Schools.....</b>	<b>10</b>
<i>Hao Xiaomei</i>	
<b>The Application of Internal Knowledge Management Tools by Using CRM Model .....</b>	<b>13</b>
<i>Fangyu Zhou, Yutong Zhou, Xi Xie</i>	
<b>Research on visual Design of Cantonese Opera Modeling elements.....</b>	<b>15</b>
<i>Yu Shengnan</i>	
<b>Study On Effective Classroom Training of Machining Specialty .....</b>	<b>17</b>
<i>Chen Liang</i>	
<b>Reconstruction design of environmental facilities in old communities in Dongguan based on Aging Society.....</b>	<b>20</b>
<i>Liang Zhu</i>	
<b>Discussion on Life Cycle Management System of Power Grid Equipment Based on Internet of Things .</b>	<b>23</b>
<i>Zhenyuan Zhang</i>	
<b>A Method of Direct Control Design Based on The Least Squares Iterative Algorithm.....</b>	<b>26</b>
<i>Jing Wang</i>	
<b>Prediction of Fungi's Decomposition Rate Based on Markov Chain and Gaussian Distribution .....</b>	<b>29</b>
<i>Hao Ouyang, Zi-Nan Li, Jia-Ming Zhu</i>	
<b>Evaluation Model of Minors' mental Health Based on Improved Factor Analysis.....</b>	<b>38</b>
<i>Xintao Cui, Jianglong Mao, Junfang Zhang</i>	
<b>Credit Rating Classification Model Based on AHP-EWM.....</b>	<b>42</b>
<i>Yaxun Dai, Bingbing Shi, Fanjie Jin</i>	
<b>Damping Attenuation Characteristics of Valve-Controlled Damping Adjustable Shock Absorber .....</b>	<b>47</b>
<i>MeiXiang Zhai, KeMing Chen, Hao Lin, XiaoJun Men</i>	
<b>Research On Dynamic Programming Model of Ordering and Transportation of Raw Materials Based on Production Enterprises.....</b>	<b>52</b>
<i>Xiaoxiao Dong, Haowei Yan, Liya Chen</i>	
<b>The Design and Management Model of Airport Taxi Based on Goal Planning .....</b>	<b>56</b>
<i>Zehui Dou, Yumei Gong, Renfang Wang</i>	
<b>Analysis Of Pollutant Particle Distribution and Content Based on Diffusion Model .....</b>	<b>60</b>
<i>Hongwei Gao, Xin Meng, Xiyan Meng</i>	
<b>The Catalytic Process of C4 Olefin Preparation from Ethanol was Investigated based on Particle Swarm Optimization .....</b>	<b>64</b>
<i>Zixuan Ge, Zhenbin Zhang, Bingbing Shi</i>	
<b>An analysis of the influence of American presidential candidates on American economy.....</b>	<b>68</b>

*Xueying Jing, Zehui Dou, Bingbing Shi*

<b>Food System Committed to Fair and Sustainable Development.....</b>	<b>74</b>
<i>Xichang Li, Shihao Lu, Kunlong Li</i>	
<b>Evaluation of Mental Health Status of Minors Based on Improved Analytic Hierarchy Process .....</b>	<b>79</b>
<i>Kun Liu, Shuo Zhang, Xichang Li</i>	
<b>Value Analysis of Users' Consumption Behavior Based on Stepwise Logistic Regression Model.....</b>	<b>83</b>
<i>Yiyao Liu, Sijia Zhang, Qian Liu</i>	
<b>NCFA: Networks Correlations Forecast and Assessment of Music Evolution.....</b>	<b>86</b>
<i>Yibo Wang, Yunjiang Han, Yi Zhu</i>	
<b>Catalytic Experimental Study on Preparation of C4 Olefin from Ethanol.....</b>	<b>91</b>
<i>Bingbing Shi, Zihang Chen, Yingyi Xie</i>	
<b>Airport Taxi Driver Decision-Making and Taxi Management Model .....</b>	<b>95</b>
<i>Renfang Wang, Xueying Jing, Zehui Dou</i>	
<b>Air-Conditioning Model for Car Sharing Based On Kmeans-Dijkstra .....</b>	<b>99</b>
<i>Renzhuo Wang, Yi Zhu, Yibo Wang</i>	
<b>Credit Decision Optimization Model Based On Linear Programming .....</b>	<b>104</b>
<i>Sunyufei Wang</i>	
<b>Analysis and Description of Oxygen Saturation based on Pearson Correlation Coefficient and Multiple Minear Regression .....</b>	<b>108</b>
<i>Yunjiang Han, Yibo Wang, Yi Zhu</i>	
<b>Analysis of C4 Olefins Prepared by Ethanol Coupling.....</b>	<b>112</b>
<i>Ziao Xing, Jincan Wang, Wenhua Li</i>	
<b>Optimization Analysis of C4 Olefin Preparation by Ethanol Coupling .....</b>	<b>115</b>
<i>Yiming Xu, Haoyu Zhou, Tengda Zhang</i>	
<b>Research And Development of Suitable Aging Platform Based on Middle-Aged And Elderly Online Shopping Market Research .....</b>	<b>120</b>
<i>Haowei Yan, Xiaoru Huo, Jiajia Liu</i>	
<b>Evaluation of Higher Education System based on index Dimension Reduction.....</b>	<b>123</b>
<i>Mengyuan Yang, Zihang Chen, Bingbing Shi</i>	
<b>Research on Value Characteristics of Consumer Behavior Based on Factor Analysis.....</b>	<b>125</b>
<i>Ziyang Yu, Xiaodie Zhuang, Jing Liu</i>	
<b>Photoelectric Intelligent Garbage Sorting Vehicle.....</b>	<b>129</b>
<i>Tengda Zhang, Jing Liang, Xiangliang Li</i>	
<b>The Catalytic Process of C4 Olefins Prepared from Ethanol was Investigated based on Multiple linear Regression .....</b>	<b>134</b>
<i>Zhenbin Zhang, Bingbing Shi, Haoran Bai</i>	
<b>Research on Converter Flue Gas Analysis Based on Penalty Factor Improved Machine Learning .....</b>	<b>138</b>
<i>Yi Zhu, Yibo Wang, Yunjiang Han</i>	

# A Survey of Human Pose Estimation Based on Lightweight Neural Network

Kuo Fu, Xia Liu, Qiyan Yan

Department of Mechanical and Electrical Engineering, Guangdong University of Science & Technology, Dongguan 523083, China

**Abstract:** With the rapid development and the popularization application of lightweight artificial intelligence technology, has deepened the communication between human beings and Information Physics in various fields. How can lightweight artificial intelligence better express human pose characteristics will maximize the work efficiency between people and things. In this paper, the lightweight intelligent system is given to human pose to express will more effectively, and the lightweight convolution neural network is used to extract human pose and other patterns. Forming a lightweight, miniaturized and stable intelligent algorithm has important theoretical significance for the application of human pose and education action.

**Keywords:** Human Pose Estimation; Neural Network; Lightweight

## 1. INTRODUCTION

Lightweight is not only the core of the future application of artificial intelligence, but also the key point to solve the popularization and promotion of artificial intelligence development planning. These problems include the continuous strengthening of vision technology, algorithms, computing power, databases, chips and other aspects. With the great development of functions, the software and hardware are becoming more and more complex and huge, the computational complexity of artificial intelligence algorithms is rising sharply, the energy consumption cost of computational neural networks is becoming higher and higher, and the amount of data is growing explosively. These have become constraints on the future development of regional artificial intelligence, How to adapt the application of artificial intelligence and its computing to miniaturization and lightweight has become an urgent problem to be solved.

The development plan for a new generation of artificial intelligence and the innovative action for artificial intelligence in high-level application education will promote the students of high-level application schools to enter the learning mode of "artificial intelligence +". At the same time, it will promote the better auxiliary learning of lightweight artificial intelligence and form the popularization of Intelligent Technology application, Therefore, this paper puts forward the research on human pose and action based on lightweight neural network, which is suitable for the current development trend.

## 2. RELATED WORKS

Lightweight artificial intelligence can improve the efficiency of chips, platforms and algorithms, and realize low-power artificial intelligence training and application

deployment in consideration of more miniaturized physical space, which can explore the specific application progress of human pose estimation. Based on the development of lightweight artificial intelligence in different dimensions, this paper explores and studies the trend in all aspects of development planning.

At the national level, the idea of innovation driven development in the 14th five year plan and the long-term goal of 2035 aims at cutting-edge fields such as artificial intelligence, quantum information, integrated circuits, life and health, brain science and so on. At the regional level, the outline of the development plan of Guangdong, Hong Kong and Macao Dawan district plans to build a smart city cluster, promote the pilot demonstration of new smart cities, and vigorously develop smart transportation, smart energy, smart municipal administration and smart community. The development plan for a new generation of artificial intelligence in Guangdong Province clearly promotes the in-depth integrated development of artificial intelligence and economy, society and industry. Dongguan issued the development plan of key emerging industries in Dongguan (2018-2025) and the development plan of new generation artificial intelligence in Dongguan (2019-2030), proposing to focus on five emerging fields and break through ten key industries. Research on the new generation of information technology and intelligent manufacturing has become the main driving point, breaking through the industry, and taking the new generation of artificial intelligence as the new driving force for future development. Almost every once in a while, we can see exciting planning or technological breakthroughs in relevant fields. The smaller the range of extreme lightweight, the collaborative lightweight of software and hardware is formed, and the lightweight AI application scenario breaks through the limit of small intelligence. Facing the popularization and application of lightweight artificial intelligence technology, it is an inevitable trend of the development of the current era.

Lightweight neural network, as the current technology of artificial intelligence research, has developed very rapidly, mainly in two aspects: one is the network structure design, the other is the compression of neural network. In terms of deep neural network: in order to obtain better performance, the data volume of lightweight neural network is increasing, so that CNN goes out of the laboratory and is more widely used in mobile terminals. There are models including Mobilenet, Shufflenet, Xception, etc. these models adopt 1x1 convolution core by improving the size of lightweight core; The idea of 1x1 convolution kernel is widely used in WaveNet proposed

by Google Deepmind in 2017 for text to speech generation, research scientists of OpenAI in 2018 proposed GLOWNet based on stream generation model to generate real pictures, and NVIDIA researchers proposed a combination of WaveGlow training neural network to generate high-quality speech.

Other research paths to overcome the technical defects of lightweight neural network CNN: in October 2017, Geoffrey Hinton published a paper entitled dynamic routing between capsules, which proposed that CapsNet is a new and improved lightweight neural network, which is represented by vector substitution mode, and the vector length represents the estimated probability of whether an object exists or not. It solves the recognition errors of CNN in the change of puzzle position or structural pattern, and has a good recognition effect on the recognition of overlapping patterns. Combined with the current CNN research hotspots and technical innovation methods, this study uses the improved CNN structure to strengthen the classification method in pose and action, so as to lightweight the role of artificial intelligence in human pose image mode.

In recent years, the lightweight neural network Yolo series has developed rapidly. Recently, PPYOLO tiny, which surpasses YOLOv5, has realized ultra-lightweight algorithm with a capacity of only 1.3M, which can quickly realize the target detection algorithm. In the current development and application scenarios of a large number of terminals such as mobile Internet, Internet of things and Internet of vehicles, the demand for deploying ultra-light human pose estimation directly on edge devices has increased sharply. Including deploying human pose estimation algorithms on raspberry pie, FPGA, DSP and other chips with extremely low hardware cost for automatic production. Our commonly used mobile app also needs to directly adopt a deep learning algorithm of no more than 6M at the terminal. The emergence of ultra-lightweight algorithm accelerates the popularization of human pose detection of small and micro electronic products.

### **3. DEVELOPMENT OF HUMAN POSE ESTIMATION**

Cong Leng, artificial intelligence chip Innovation Research Institute of Chinese Academy of Sciences, published an article in China Science Daily. The second half of intelligence: the rise of lightweight artificial intelligence; It is pointed out that tiny AI will push artificial intelligence to the mainstream and turn AI from a high threshold technology giant competition to an intelligent ecology that benefits people's livelihood. MIT science and technology review also listed "tiny AI" as "the world's top ten breakthrough technologies" in 2020. In its selection reasons, it wrote: "Lightweight intelligence makes the existing services such as voice assistant and mobile camera better and faster. Therefore, lightweight artificial intelligence will also enable human pose estimation to be applied to more scenes, such as learning situation analysis based on classroom scene and classroom virtual teacher assistant requiring faster response time<sup>[4]</sup>.

In terms of model building, the application of human pose ACADEMIC PUBLISHING HOUSE

estimation and behavior recognition based on visual information has always been a hot research and mature technology<sup>[5]</sup>. The application of lightweight neural network to human pose estimation has developed rapidly in recent years, such as the lightweight behavior recognition method based on key frame. A lightweight convolution neural network is designed as a feature extraction network, which can make efficient use of the apparent information and motion information of video<sup>[6]</sup>. This research idea can consider the real-time human pose estimation based on lightweight network. According to the requirements of mobile devices for algorithm parameters and computation, a real-time key point detection algorithm based on lightweight network can be established in reference<sup>[7]</sup>. In the process of adaptive visual decision-making, teachers' behavior recognition should be considered, and teachers should be tracked and predicted at the same time. In reference<sup>[8]</sup>, the deep enhancement network of behavior decision-making is used for visual tracking. Similarly, the lightweight network based on behavior decision-making can be used for visual information processing.

In the implementation of virtual lightweight intelligent teacher assistant, there is less research on specific applications, mostly based on model and theoretical research. Literature<sup>[9]</sup> built a virtual action tutor for action online learning, and the learning behavior is more extensive. Literature<sup>[10]</sup> analyzes the emotional performance of virtual teachers, designs the behavior expression mode of virtual teachers from four aspects: facial expression, body pose, voice and speech, and discusses the application advantages and value of virtual teachers based on emotional calculation combined with application cases, so as to provide reference for the research and practice of virtual teachers. Literature<sup>[11]</sup> established a framework for intelligent evaluation of teachers' teaching behavior through teaching video. Literature<sup>[12]</sup> intelligent classroom assistant adopts natural language processing and emotion analysis technology to naturally interact with students to realize interesting learning and personalized teaching. It puts forward the teaching framework of intelligent classroom assistant and realizes the specific application of classroom teaching. The construction of virtual teacher assistant with emotional expression and formulate teaching process application strategy assistant, so as to assist teachers in effective emotional interaction with learners and lightweight artificial intelligence devices.

### **4. SUMMARY**

As a popular key technology, lightweight artificial intelligence will change or even subvert many traditional industries, such as education, health and health, industrial automation and transportation. Lightweight neural network will also create some new products, models and even industries. Facing the subversive change, we should actively embrace the change and explore the problem of human pose estimation in teaching. Build a virtual teaching assistant with lightweight artificial intelligence miniaturization and micro collaborative assistance. Combined with the practice of lightweight artificial

intelligence, human pose estimation is applied to break the boundary between learning and teaching, and explore the relevant research progress. Pave the way for further lightweight original innovation and more subtle fields, and realize ultra light fields such as intelligent assistant.

## REFERENCES

- [1] New generation artificial intelligence development plan [z]. Scientific and technological innovation and productivity, 2017 (4): 52-66.
- [2] Action plan for artificial intelligence innovation in Colleges and universities [z]. China information technology education, 2018 (8).
- [3] Jin Chen, Wenjing Lv. Artificial intelligence and new engineering talent training: a major turn [J]. Research on higher engineering education, 2017 (6): 18-23.
- [4] Pan Yunhe. Artificial intelligence 2.0 and the development of Education [J]. China distance education, 2018 (5).
- [5] Ls A, Gang Y B, Jy A, et al. Human Pose Estimation and Its Application to Action Recognition: A Survey[J]. Journal of Visual Communication and Image Representation, 2021.
- [6] Zhou Yuxin, Bai Hongyang, Li Wei, Guo Hongwei, Xu Xiaokang. Research on lightweight behavior recognition method based on key frame [J]. Journal of instrumentation, 2020, 41 (07): 196-204.
- [7] Hu JiangHao, Wang Hongyu, Qiao Wenchao, Ma Jingxuan. Real time human pose estimation method based on lightweight network [J / OL]. Computer Engineering: 1-11 [2021-05-26].
- [8] Yun S, et al. Action-Decision Networks for Visual Tracking with Deep Reinforcement Learning[C], IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2017.
- [9] Hsin-Hung Chiang, Wei-Ming Chen, Han-Chieh Chao, De-Li Tsai. A virtual tutor movement learning system in eLearning[J]. Springer US, 2019, 78(4).
- [10] Zhu Ke, Zhang Siyan, Liu Mengyu. Design and application advantages of virtual teacher model based on emotional computing [J]. Modern educational technology, 2020, 30 (6): 78-85.
- [11] Zheng Yuhuang. An evaluation method of teachers' teaching behavior based on pose recognition [J]. Software engineering, 2021, 24 (04): 6-9.
- [12] Liu Yanyan, Huang shuanggen, Huang Fuyu. Intelligent classroom assistant: teaching tools in the context of artificial intelligence [J]. Software guide, 2019, 18 (09): 217-220.

# The Trinity of Government School and Enterprise Advantages and Elements Sharing Collaborative Innovation Center

*Ding Wenhui*

*School of Management, Guangdong University of Science & Technology, Dongguan, China*

**Abstract-**Nowadays, due to the continuous transformation of the economic structure, new adjustments to the training programs of higher education institutions are constantly being proposed, and we need to provide more ideas.In particular, the teaching of colleges and universities should be aligned with the training goals of application-oriented and skill-oriented.Therefore, the "trinity of government, school and enterprise" has become an important way out for the teaching reform in colleges and universities.So how can we effectively achieve this goal? Therefore, we can try to build a shared collaborative innovation center based on the trinity of government, university, and enterprise advantages. This article specifically focuses on this topic to give specific explanations.

**Keywords-**Higher Education Reform; The Trinity Of Government; School And Enterprise; Advantage Elements; Sharing; Collaborative Innovation Center

## INTRODUCTION

Unlike the past simplification of the higher education model that focused on theoretical talents as the main training target, the reform of higher education now requires us to achieve hierarchical teaching, and for some undergraduate colleges and universities, we must integrate with the training goals of applied and skill-based training.This is also an important aspect of the structural reform on the supply side of higher education.Deepening innovation and entrepreneurship education is an important starting point for the supply-side structural reform of talent training in applied undergraduate colleges[1].And what are the specific aspects of deepening innovation and entrepreneurship education? So we need to explore the emerging model of "the trinity of government, school and enterprise".In this regard, we already have some more successful cases.

## THE NECESSITY OF A COLLABORATIVE INNOVATION CENTER FOR SHARING THE ADVANTAGES OF THE TRINITY OF GOVERNMENT, SCHOOL AND ENTERPRISE

Today, our school education is mainly the teaching materials for the discipline, and most of the content of the disciplines is "past form".This is the crystallization of our predecessors to nature and society, and solidify in the textbooks learned today.However, life experience tells us that many of our knowledge and abilities are not learned in the books written in the former person, but learned in life practice.The students we have to cultivate today are "future" talents, so only the knowledge of the past, let students deal with infinite changes, these knowledge is far less than enough.Therefore, today we are thinking about

educational reforms, it is necessary to pay attention to the civilized achievements of the inheritance, but also pay attention to the special requirements of future society on talents.To achieve a collaborative innovation center for sharing the advantages of the trinity of government, school and enterprise, it is necessary to take school-enterprise cooperation as the entry point[2].Then focus on strengthening the construction of the platform to create a perfect model of collaborative training.The necessity of this model is reflected in the specific requirements for the application-oriented undergraduate training mechanism.Let's discuss its necessity in detail below.

### Conducive to cultivating innovative talents

Innovation is a senior performance of human subjective energy, it is the original ability of human unique creation and development, which is the eternal motivation to promote social progress and development.Because innovation is the source of knowledge evolution.Knowledge is limited, innovation is unlimited.Only innovation can drive all the development in the world, thereby driving all progress.If a nation or a country wants to advance to the forefront of The Times and become a world power, it cannot do without innovative thinking, still less all kinds of innovation.At the same time, countries around the world have also increasing the investment and intensity of cultivating and researching innovative talents in various aspects.Cultivating innovative talents is one of the greatest demands for talent cultivation in our age.It can be said that this goal is also in line with the needs of the reform of the talent supply side.Especially in the new era, there are also new requirements for the cultivation of innovative talents.Innovative talents in the new era must be talents who have received comprehensive training in science, technology, economy, and culture[3].They must be talents who can meet the development needs of the industry in their own quality.Therefore, through "the trinity of government, school and enterprise" model, education, talents, industry and other chains can be effectively combined, so that the cultivation of talents can achieve the school and society and other aspects of the multiple influence, so as to significantly improve the innovation ability of talents.

### Conducive to cultivating talents who can meet business model changes

This is an era of highly developed informatization, as well as an era of 5g commercial and Internet+.This has brought many outlets and opportunities to the transformation of business models[4].This also requires us to step out of the traditional business model, create a new business model

based on Internet +, and integrate various online and offline resources to meet the needs of 5g commercial use. At this time, we are required to cultivate talents so that they have the development and operation capabilities of the front, middle and back ends. And this is also one of the more lacking areas in our talent training today. Therefore, through the "the trinity of government, school and enterprise" model, 5G commercial and Internet + can be brought into the talent training, so as to effectively cultivate talents who can meet the business model reform, and then let them participate in the innovation and innovation, and finally make the quality of talents more in line with the times[5].

Meet the specific needs of applied undergraduate training In order to establish a modern vocational education system, the Ministry of Education launched the transformation and reform of colleges and universities in 2014, requiring the transformation pilot colleges and universities to construct a governance structure with the participation of industries and enterprises, take technology application as the guidance, highlight the collaboration between industry and education, and establish and improve a comprehensive, innovative and application-oriented talent training model to meet the market demand. In this context, based on the guiding ideology of the integration of production and education, we will carry out in-depth innovation and entrepreneurship education for college students, establish an educational platform for off-campus college students' dual innovation practice, integrate the tripartite resources of school, enterprise and government, improve the collaborative education mechanism, cultivate high-quality technical application-oriented talents with innovative entrepreneurial spirit and practical ability, it has become one of the important contents in the reform of talent training mode in application-oriented universities.

#### THE MAIN PROBLEMS IN THE COLLABORATIVE INNOVATION CENTER OF THE TRINITY OF GOVERNMENT, SCHOOL AND ENTERPRISE SHARING ADVANTAGES AND ELEMENTS

The training mode is relatively simple

The current application-oriented undergraduate colleges mainly adopt the first classroom on campus, in other words, courses on innovation and entrepreneurship and the second classroom on campus will be set up in the talent training program, teachers guide students to participate in various innovation and entrepreneurship competitions, carrying out innovation and entrepreneurship education for college students lacks the participation of the third classroom outside the school, that is enterprises, industries and local governments[6]. This training mode leads to a single form of innovation and entrepreneurship education and a short period of effect.

The disconnect between classroom training and practice training

Some applied undergraduate colleges have unclear understanding of the significance of innovation and entrepreneurship education for college students, focusing on innovation and entrepreneurship, and lightly on the essence of education[7]. It will continue to innovative entrepreneurship education as to a few innovative,

entrepreneurial intention of students in education, that is one of the few people's patent, one-sided pursuit of students' office and financing amount, no innovation entrepreneurship education into the whole talent cultivation system, no form for the development of innovative business knowledge, professional knowledge and industry coupling interaction.

There is a distance between institutions and industries The division of disciplines, majors, and knowledge is common in applied undergraduate colleges. In addition, the "wall" between schools and industries is also a natural barrier to achieve the goals of innovation and entrepreneurship education. How to bridge the boundaries of disciplines and industries? There is an urgent need to explore the establishment of an innovation and entrepreneurship education platform that crosses the boundaries of industry and academia.

#### Insufficient protection system

Schools, enterprises, and the government often lack long-term cooperation mechanisms due to unclear responsibilities and rights, especially the lack of motivation for enterprises to jointly cultivate innovative and entrepreneurial talents.

#### SOLUTIONS TO MAJOR PROBLEMS

##### Broaden the channels for talent training

The current demand for talent is very large, but there is still a big gap in talent supply and demand. Therefore, to better train talents, we need to broaden talent training efforts. Not only to strengthen talent cultivation in simple learning, but also to build more stages for talent cultivation. Only by allowing them to expand their knowledge in more places and discover their shortcomings can they pay more attention to their shortcomings and make more efforts in self-reinforcement. Of course, in order to better cultivate talents, we also need to give play to the role of schools and give more space for talents to grow, so as to cultivate talents more in line with the needs of social development. Integrate off-campus resources to build a third classroom, and carry out various off-campus double innovation practice projects. For example, schools and enterprises jointly build youth innovation spaces, guide students to participate in various innovation and entrepreneurship competitions, help businesses incubate promising projects, and provide a free entrepreneurial platform for university students. Hiring successful business people as off-campus mentors to provide students with guidance on career planning, internship practice, entrepreneurship and employment[8]. It can also design off-campus dual innovation practice projects for college students around smart agriculture, rural e-commerce, beautiful villages, targeted poverty alleviation and other national construction projects, and optimize the professional group dual innovation curriculum system and practical teaching content.

##### Achieve multiple integration of training models

Realize the joint development of school and enterprise, compile a talent training plan for entrepreneurship and innovation, curriculum syllabus and handouts for entrepreneurship and innovation practice, school-enterprise jointly completed teaching reform projects and

published papers. By transferring innovative practical courses to enterprises to carry out "classroom revolution", schools and enterprises can jointly teach entrepreneurship practical courses, jointly guide students to write entrepreneurship practical course reports, and jointly reform the whole process of practice assessment and evaluation. Through the method of "on-campus tutor + off-campus tutor", students are instructed in their graduation design, so that innovation and entrepreneurship education runs through the whole process of talent training, and its integration with professional education is realized. At the same time, through the local government and enterprises jointly held the county brand release conference, students to promote products on site, stimulate students' innovation consciousness. Then, through the rural e-commerce poverty alleviation lecturer project and the school-enterprise cooperation, students are guided to make the bidding documents for the operation of rural e-commerce service center at the county level, and build the county agricultural products traceability system platform.

Realize the real synergy between universities and industry. The school and the enterprise jointly declare and build a provincial-level university student off-campus innovation and entrepreneurship practice education base, which can not only help enterprises to better understand the needs of local industrial development for the ability of entrepreneurial talents, but also enable corporate mentors to enter the school and in-school mentors to enter the enterprise. In this way, it is possible to form a faculty of

Table 1. MAIN PROBLEMS AND SOLUTIONS

main problem	Single training mode	The classroom is out of touch with practice	There is a distance between institutions and industries	Insufficient protection system
solution	Broaden training channels	Achieve the integration of training models	Realize the real synergy between the two	Establish a complete guarantee system

The basic structure of the collaborative innovation center for sharing the advantages of the trinity of government, school and enterprise is shown in the figure1:

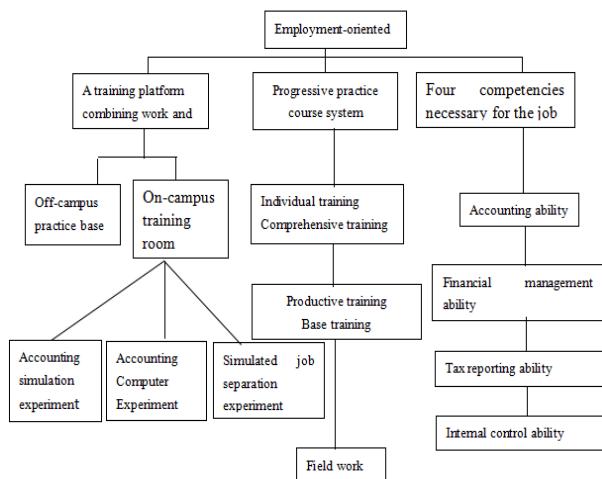


Figure 1. Basic structure diagram

## CONCLUSION

We will improve personnel services and speed up personnel training. The cultivation of talents is a systematic project, which needs universities to play an active role. Universities and institutions across the country

innovative and innovative education teachers with "specialized creation, financing and integration + combination of expertise" and achieve deep integration of production and education. At the same time, supported by the construction of the base platform and relying on the resources of the three parties of the university, through carrying out various forms of entrepreneurship and innovation practice projects, further achievements have been made in the training mode of entrepreneurship and innovation talents, courses, teaching materials, teaching team, educational reform and research, assessment and evaluation, organization and management, rules and regulations, and integration of industry and education.

## Establish a complete guarantee system

From the perspective of government-school-enterprise-society, systematically build out-of-school practical education guarantee conditions for the cultivation of entrepreneurial and applied talents. The government provides policy support for innovation and innovation education, and encourages universities to establish unique innovation and innovation bases. The school integrates entrepreneurship and innovation education with professional education, strengthens the practical education of entrepreneurship and innovation, improves the management rules and regulations of innovation and entrepreneurship outside the school, as well as the construction of assessment and evaluation incentive mechanism, and provides the corresponding fund guarantee. The main problems and solutions are shown in

Table 1. MAIN PROBLEMS AND SOLUTIONS

provide high-level talent training programs, and provide convenient services in recruitment, talent introduction and project incubation. Obviously, better service for talents can speed up the training of talents, enhance the strength of talents, so that talents can better serve the society. The trinity of government, school and enterprise share the advantages of collaborative innovation center, which is an adjustment and reform of application-oriented undergraduate education mode under the new era situation. For this reason, we need to reflect on the original training model, constantly focus on the deficiencies in the ghost zone, and make relevant adjustments to the training model.

## ACKNOWLEDGEMENT

Project of Guangdong University of Science and Technology in 2019: Collaborative Innovation Center for Sharing Advantages of Government, University and Enterprise (GKY-2019CQYJ-7).

## REFERENCES

- [1] Yang Lei; Han Daliu. "Project-driven, base support, school-enterprise-government trinity" off-campus innovation and entrepreneurship education model construction[J]. Modern Business and Trade Industry, 2021, 2(8): 89-90.

- [2] Dong Yanmao, Ye Changqing et al. Trinity tutorial system. The idea of college-enterprise cooperation to cultivate college students' innovation and entrepreneurship ability [J]. Higher Education of Chemical Industry, 2020, 4 (28) : 81-82.
- [3] Chang Xiaoyu. Innovation and entrepreneurship education and teaching reform in higher vocational colleges based on the integration of industry and education [J]. Education and occupation. 2019(21):76-80.
- [4] Yang Lei, Wang Jin, Wang Zhuyun. Exploration on the Construction Content of Cross-border E-commerce Innovation and Entrepreneurship Practice Base of International Trade Major -- Taking Xi 'an Siyuan University as an Example [J]. Modern Business Industry. 2018(31):45-46.
- [5] Yang Jingang, Yang Ming. Construction and Operation of Entrepreneurship Education System in Colleges and Universities Based on Collaborative Innovation between Universities and Enterprises[J]. Theoretical research and practice of innovation and entrepreneurship. 2018(12):53-54.
- [6] Tang Yu, Yu Juan, Wang Bing, Wei Jinmin. Exploration and practice of the talent training model of "integration of industry and education, school-enterprise cooperation" under the background of local universities' transformation [J]. College Education. 2020(06):160-163.
- [7] Exploration on the Cultivation of Postgraduate Scientific and Technological Talents under the Pattern of High-quality Development -- Comment on Research on the Cultivation of Innovative Scientific and Technological Talents Based on the Quality Model [J]. Zhou Jiadong. Science and Technology Management Research. 2020(08):278.
- [8] Chang Jie, Qiao Bin. Performance Evaluation of Industry-University-Research Collaborative Innovation in S&T Small and Medium-sized Enterprises [J]. Statistics and Decision. 2020(06):185-188.

# Study on the Influences of Application Colleges and Universities Teachers on Ten Education System Construction

Zhang Yijun, Yuan Dan

Guangdong Institute of Science and Technology, Dongguan, Guangdong 523083, China

**Abstract:** Colleges and universities are the place to cultivate people, the base of educating people for the party and the country, and the cradle to cultivate young people with “four services” consciousness and “four correct understandings” in the new era. Teachers in application colleges and universities play an important role in the process of effectively constructing the “ten education system”. In the new era, teachers in application colleges and universities should be guided by the socialist core values, aim to serve the local economic and social development, fully excavate and exert the education function of curriculum, scientific research, practice, culture, network and psychological work, support the improvement of educational mechanisms in management, service, funding and organizations, to build a comprehensive, systematic and long-term education mechanism.

**Key words:** Ten Education System; Application Colleges And Universities; Teachers; The New Era

## 1. THE IMPORTANCE ANALYSIS OF THE COLLEGE AND UNIVERSITY TEACHERS IN UNDERTAKING THE MISSION OF EDUCATING PEOPLE

Application colleges and universities are application-oriented, the purpose of teaching and scientific research activities is to serve the development of local economy, society and industry. Cultivating a first-class teacher team is an important way to promote the transformation and development of application colleges and universities. The construction of teacher team should be based on the cohesion of team strength of each teacher based on their own work. Under the guidance of Xi Jinping Thought on Socialism with Chinese Characteristics for a New Era, teachers should conscientiously study and implement the spirit of relevant conferences on school construction and student development, deepen the spiritual implementation, and take on the mission of educating people. In this regard, under the background of the new era, it is of great theoretical and practical significance to explore the mechanism of the college and university teachers in the construction of “ten education system”, give full play to the innovation of comprehensive education and improve the quality of talent training.

*The Implementation Outline of the Quality Improvement Project of Ideological and Political Work in Colleges and Universities* presented the concept of “ten education system”, stressed the importance of educational system in curriculum, scientific research, practice, culture, network,

psychological work, management, service, funding and organizations<sup>[1]</sup>. According to document guidelines, we explore the importance and mechanism of teachers in the construction of “ten education system”, focus on the analysis of the role of teachers in the curriculum, scientific research and practice education system construction.

Teachers should take moral education as the foundation and ideal education as the core, focus on socialist core values-led, and improve the long-term mechanism of systematic education. In order to form a comprehensive education pattern, and cultivate students who have all-round development of virtue, intelligence, physical conditions, mind status and community service in the new era. We should effectively build an all-round, systematic and long-term education mechanism in the new era, so that the talents we cultivate can participate in the great responsibility of socialist construction and national rejuvenation.

## 2. THE CHARACTERISTIC ANALYSIS OF THE TEACHERS IN APPLICATION COLLEGES AND UNIVERSITIES

### 2.1. Great responsibility

Teachers in Application colleges and universities play a certain exemplary role in teaching, scientific research, practice, culture, network, psychological work and other works, take an important role to guide the way of students' growth. Teachers' self-quality and political accomplishment are closely related to the talent training, teaching and research, cultural inheritance and service innovation of colleges and universities.

### 2.2. Application-oriented

The college and university research team focuses on the theoretical and applied research work in a certain field through different cooperation modes, and undertakes the regular tasks of cultivating students. As the application colleges and universities are application-oriented, the purpose of teaching and scientific research activities is to serve the development of local economy, society and industry. The research work of scientific research teams in application colleges and universities are mostly focused on application-oriented research. The teachers play a very important role in the promoting of transformation and application of achievements.

### 2.3. Playing an important role in innovation and entrepreneurship support system

The raise of “Mass Entrepreneurship and Innovation” strategy is to stimulate the vitality of entrepreneurship and innovation. On Oct 12th 2021, the General Office of the State Council issued *the Guiding Opinions on Further*

*Supporting the Innovation and Entrepreneurship of Higher Schools Students ("Opinions").* The Opinions pointed out that college and university students are the new force for mass entrepreneurship and innovation, and it is of great significance to support them in innovation and entrepreneurship. As a place where high-quality talents gather, the college and university teachers' team plays an important role in the whole process of further supporting students' innovation and entrepreneurship. And, this is especially important to the teachers in Application colleges and universities.

### **3. THE STRATEGIC ANALYSIS OF APPLICATION COLLEGES AND UNIVERSITIES TEACHERS IMPROVING "TEN EDUCATION SYSTEM" CONSTRUCTION**

Application colleges and universities teachers play an important role in the construction and improvement of "ten education system", especially in the curriculum, scientific research and practice education system. In this article, we mainly introduce the role of teachers in the construction of these three aspects and how do they promote the construction.

#### **3.1. Deepening the reform of "Curriculum Ideological and Political", improving the curriculum education system**

Ideological and political construction in colleges and universities is needed by the times, majors and talents. The curriculum ideological and political construction needs to realize the new dimension of teaching objectives, new depth of teaching content, new perspective of teaching methods and new height of teaching effect, so as to achieve the fundamental goal of talent training, discipline construction and high-quality and all-round development of students<sup>[2]</sup>. Teachers should give full play to the main role of curriculum education, constantly innovate in improving classroom teaching design, make full use of the ideological and political education function carried by the curriculum, and unify curriculum ideological and political education and professional knowledge education in the new era.

#### **3.2. Strengthening innovation and entrepreneurship education, innovating the scientific research education system**

The reform of innovation and entrepreneurship education is an urgent practical need for colleges and universities to cultivate high-quality talents. We should deepen the reform of innovation and entrepreneurship education in colleges and universities, run through the whole process of talent training, so as to achieve more sufficient and higher quality employment for students<sup>[3]</sup>. Colleges and universities should strengthen the training of teachers' teaching ability and quality of innovation and entrepreneurship education, improve the curriculum system, and further promote the reform of practical teaching. Teachers should innovate teaching design, promote the organic integration of professional

knowledge education and innovation and entrepreneurship education, encourage students to establish innovation and entrepreneurship associations, lead them to participate in competitions, promote the transformation of scientific research and teaching achievements, and thus improve the quality of talent training.

#### **3.3. Adhering to the combination of theory and practice, improving the practice education system**

In addition to promoting the improvement and innovation of the practice education system with the school as the main body, teachers should also actively integrate the practical resources around them, relying on high-tech zones, industrial parks, enterprises and institutions, patriotic education places, etc., to create conditions for the expansion of diversified practice platforms and entrepreneurship bases. On top of that, We should continue to enrich the practice content, innovate practical forms, and help students to carry out extensive social practice activities such as social research, volunteer services, work-study program and scientific and technological inventions.

### **4.CONCLUSION**

Teachers in application colleges and universities should adhere to the "student-centered", follow the regulations of ideological work, teaching and student growth, give full play to the curriculum, scientific research, practice, culture, network, and psychological education function, improve the management, service, funding, organization education mechanism, to promote "then education system", contribute our own strength in cultivating the youth who have "four service" consciousness and "four correct understanding" of the new era.

### **ACKNOWLEDGEMENTS**

This paper is funded by the Party Building project of Guangdong University Party Building Research Association (Project No.: 2020MB041).

### **REFERENCE**

- [1] The Party Leadership Group of the CPC Ministry of Education. The Implementation Outline of the Quality Improvement Project of Ideological and Political Work in Colleges and Universities[EB/OL][2017-12-06].[http://www.moe.gov.cn/srcsite/A12/s7060/201712/t20171206\\_320698.html](http://www.moe.gov.cn/srcsite/A12/s7060/201712/t20171206_320698.html)
- [2] Zhou Lin. Research on Ideological and Political Construction of College Curriculum in the New Era [J]. Educational Theory and Practice, 2021, 41 (27): 34-36.
- [3] The General Office of the State Council. The Guiding Opinions of the General Office of the State Council on Further Supporting the Innovation and Entrepreneurship of College Students [EB/OL][2021-10-12].[http://www.gov.cn/zhengce/content/2021-10/12/content\\_5642037.htm](http://www.gov.cn/zhengce/content/2021-10/12/content_5642037.htm).

# Talking about the Application of Tea Culture in the Management of Class Moral Education in Vocational Schools

*Hao Xiaomei*

*Department of Tourism Service, Zhongshan Technician College, China*

**Abstract:** Tea culture carries a rich education function. Integrating tea culture into the moral education management of vocational schools, it can not only carry forward the tea culture, but also make the moral education work twice the result with half the effort, reflecting the leading role of tea culture and the value of education. This paper is that the author combines the practice and discusses the tea culture function of moral education in the class moral education management of vocational schools, and provides reference for improving the class moral education management of vocational schools.

**Key words:** Tea Culture Vocational School; Class Moral Education; Management And Education

## I. THE DEVELOPMENT HISTORY OF TEA CULTURE

Our confidence in traditional Chinese culture comes from the understanding and recognition of traditional Chinese culture, and comes from the understanding and recognition of traditional cultural ideas and value system. Tea originated in ancient China and spread in the western world. The harmonious tea culture integrates the culture and thoughts of Chinese Buddhism, Taoism and Confucianism. It is extensive and profound, rich in connotation and long history. It is one of the main representatives of Chinese traditional culture.

Tea, also known as 'Ming', 'Da', was mainly produced in Bashu in ancient China. It was widely spread from south to north and west to east. It has developed into one of the three traditional cultural drinks in China and even the world. Lu Yu in the Tang Dynasty, 'The Tea Classic Six Drink ': tea is drinking, originated in Shennong, heard in Duke Lu and Zhou'. In other words, tea was originally served as a drink from Shennong, and it spread during the time of Duke of Zhou, the first sovereign of the Zhou Dynasty. By the Tang, Song Dynasties, Ming and Qing Dynasties, tea had become an indispensable necessity for people's work and daily life (namely firewood, rice, rice, oil, salt, sauce, vinegar and tea), commonly known as the saying of "tea as food, which is no different from rice salt". To put it simply, the long history of tea culture can be seen as an epitome of the development of China's 5,000 years of civilization.

## II. THE OVERALL STRUCTURE AND CONTENT OF TEA CULTURE

Some scholars have analyzed that the overall structure of ancient Chinese tea culture and traditional Chinese medicine tradition should be "one body and two wings", that is, a whole body with Chinese Buddhism, Taoism and

Confucianism as the core, while Chinese medicine and modern tea ethics and tea ceremony are the two core wings. These two wings are the ingenious comprehensive application and successful integration of Chinese Buddhism, Taoism, Confucianism, Chinese medicine, tea ethics and tea ceremony. Traditional Chinese medicine nourishes the body, cultivates the heart with tea morals, and enjoys the tea ceremony. The body and mind are united, the value is shared, and the unity of Buddhism, Taoism and Confucianism is returned.

There is virtue in tea and Tao in art. There are three realms of tea culture. One is to drink tea: only focus on the material properties of tea, and treat tea as drinking water. The second is tea tasting: because drinking tea is not just drinking water, it is the enjoyment of beauty among various senses of sight, hearing, touch and taste, which is called tasting. The third is to gain the Tao: drinking tea is like life, tea tastes character. The famous Chinese tea expert Mr. Zhuang Wanfang proposed "Chinese tea morality", and its cultural content is mainly "incorruptibility, beauty, harmony and respect". The explanation given by Mr. Zhuang is: "Lian" means "clean, diligent and thrifty"; "Mei" means "delicious, recreational"; "Harmony" means "honor and honesty"; "Jing" means respect and help others. The tea ceremony of Cha De fully interprets the connotation and essence of Chinese traditional tea culture.

## III. THE PUBLICITY AND EDUCATION FUNCTION OF TEA CULTURE

Culture to transform people, cultural education. Chinese tea culture plays an imperceptible role in education and edification of people's quality. Cultural education and edification can not only make people cultivate good heart, but also make people cultivate good quality and cultivate good temperament. As the so-called "sneak into the night, moisten things silently" and make people invisible.

Traditional Chinese tea culture has rich connotation, from traditional handicraft facilities, to daily use of utensils; from ordinary living utensils to piano, music and calligraphy, tea culture includes it. Colleges, vocational schools is important Chinese tea culture propaganda position, through the popularization of culture and carry forward Chinese tea culture, can influence and cultivate students' ideological personality quality, the pursuit of values, character, cultural values, moral cultivation, aesthetic life, influence and cultivate students' positive psychological state, behavior, image etiquette shaping, communication ability, interpersonal ability, cooperation ability, students' professional quality and future planning

of life. By spreading Chinese tea culture and learning Chinese tea culture, you can 'nourish the heart with tea', 'respect guests with tea', 'cultivate virtue with tea', 'cultivate honesty with tea'.

#### 1. Repair heart and care

From the point of view of psychological state, vocational school students have strong self-awareness, poor psychological concentration, prone to impetuous, lack of sense of security. Learning tea art, learning tea ceremony, knowing the tea ceremony, can cultivate students 'habit of quiet learning, cultivate students' calm and relaxed mood. Because tea is mild, tea drinking can provide health care, relieve inner anxiety, and relieve the spirit of stagnation.

By learning tea culture, enhance cultural confidence, cultivate the humanistic spirit of humility and comity, tolerance and open-minded, peace and tranquility, learn to slow down, study, concentrate on work, perception and reflection, purify the heart, shape character, sublimate personality, and live out the wonderful and value of life.

#### 2. Aesthetic and educational function

Tea tea activities are suitable for all young to enjoy elegance and customs. Tea activities is not only a kind of life art, but also an artistic life. Its contents include: dance, literature, painting, books, music, clothing, flowers, etc. By learning traditional tea art, students in vocational schools can further improve their aesthetic and art qualities.

#### 3. Enhances interpersonal relationships

In the network era, vocational school students increase online communication, reduce face-to-face communication, lack of interpersonal skills, and weak teamwork ability. With tea and tea sharing can promote each other to open their hearts, communicate and gather emotions, promote value recognition, and enhance interpersonal communication ability and team integration ability. Vocational school students can learn to treat guests from tea worship tea, learn to be grateful and tolerant, respect and love each other, and coordinate interpersonal relations. Through the tea culture activities, we should pursue the harmony between the heart and the outside world, achieve a suddenly enlightened state of mind, promote the improvement and maturity of psychological quality, and cultivate a positive and healthy sunshine mentality.

#### 4. Tea ceremony education function

The etiquette of traditional Chinese tea art is a traditional ceremony and ceremony that students should strictly follow in the process of conducting traditional tea activities and tea art services. Learning the etiquette of traditional Chinese tea art requires students to master the standards of traditional Chinese tea art etiquette, tea procedures, steps, rules of tea culture etiquette; but also need students to understand the commonly used tea ceremony and different Chinese traditional tea gifts, such as bowing, tea ceremony, palm ceremony, tea ceremony and so on.

Learning the etiquette of traditional Chinese tea art will help encourage students to fully experience the traditional behavior and culture of respecting ritual in a strong sense of ceremony, and help to improve and cultivate students' quality of etiquette and the shaping of image.

### IV. APPLICATION OF TEA CULTURE IN CLASS MORAL EDUCATION MANAGEMENT

The educational training of talents cannot be separated from three kinds of educational environments: family education, school education and social education. Among them, school education plays a decisive role in the growth of talents, and the most important organization in school education is the class. The application of tea culture to the moral education management in the class can get twice the result with half the effort.

1. Tea culture is integrated into the layout of the class living environment of the school. School class living environment layout design to fully combine the Chinese tea culture, the living environment style to fully achieve natural nature, quiet and elegant, refined, plain and pure and fresh, and fully join the elements of tea culture, blackboard newspaper layout design, humanized layout, warm classroom light, wall decoration color, etc., let every object contains tea culture, every wall will speak, create a very elegant place for learning.

I served as the head teacher of a new class in September 2017, and when I arranged the class culture, I was interested in introducing tea culture. On the wall on the side of the classroom is the student style display column. Through discussing with the class cadres, we posted a picture of the ancient tea tree, selected the learning star and the star of discipline every day, and then posted their photos on each branch, so that each student can establish the consciousness of sending new buds and new tea for the big tea tree. Above the blackboard at the back of the classroom hung posters with "Jing", "Jing", "Jing", "Jing" and "He" brush characters. Make class convention: a clean environment, quiet behavior, respect teachers, get along harmoniously with students, help each other more progress every day. Let the students in the daily life are imperceptible practice of the concept of tea culture.

2. Tea culture courses should pay attention to the integration of teaching and education in tea culture courses. In the tea art service course of my class, I not only taught students the kind of tea, the nutritional value of tea and tea making technology, but also cultivated their etiquette through tea art performance. Such as: ceremonial etiquette, bowing ceremony, serving tea ceremony, etc. Students not only learned the knowledge of tea art culture, but also learn to treat guests from respecting tea and tea, learn to be grateful and tolerant, respect and love each other, humble and polite.

3. school held a morning reading report performance at the end of the semester. After discussing with the students, our class selected to recite the commentary of oolong tea art. Every morning, the students wear classical tea art clothes and read the tea art commentary in unison under the background of

elegant guzheng music. Under the influence of tea culture, they seem like integrated into a work of art, one by one like ladies and gentlemen, elegant and peaceful.

4.Has set up a tea art performance society to give full play to the educational role of the tea art cultural society.Through the establishment of school tea art performance club, can be some students interested in Chinese traditional tea culture all together, through their radiation and drive to promote the spread of tea culture, play the education function, can also through the club cultivate students' professional interest, explore talent, create reserve force for the development of tea culture.

5.Through the tea art competition, education transformation of underachiever, establish its self-confidence.A female classmate in the class was relatively withdrawn, and I encouraged and persuaded her to participate in the school tea art competition.After hard practice, her group won the third prize in the school.She found confidence from the success of the match and gained a friendship in the match.

6.Carries out the role of dormitory tea culture environment education.The concept of tea culture is integrated into the cultural construction of each dormitory, making the dormitory environment as warm and natural as home, and the relationship between the dormitory members is more harmonious.

## V. CONCLUSION

To sum up, tea culture has an important leading role in the class moral education management, so, as educators we need to clear according to the school education goal tea culture import content, promote the use of tea culture in class moral education management, relying on the tea culture for students' ideological education work, to ensure that the class can faster and better development.

## REFERENCES

- [1] Luo Xueliang, Chinese Tea Ceremony and Tea Culture, Golden Shield Press, 2016, First Edition, 5-8.
- [2] Zhang Houjun, Application of Tea Culture in Moral Education in Colleges and Universities, Fujian Tea, 2018, 2.
- [3] Wang Mengmei, On the Integration of Tea Culture Ideas and Educational Psychology, Fujian Tea, 2018, 180-181.
- [4] Huang Jin, Teaching Exploration and Thinking of Tea Art Course, Agricultural Archaeology, 2015, 234-235.
- [5] Fan Chao, Education and Research of Chinese Tea Ceremony Culture on the Cultural Inheritance of Higher Vocational Students, Fujian Tea, 2017, 206.

# The Application of Internal Knowledge Management Tools by Using CRM Model

Fangyu Zhou<sup>1</sup>, Yutong Zhou<sup>2</sup>, Xi Xie<sup>3</sup>

<sup>1</sup>Guangdong University of Science & Technology, Guangdong, 523083, China;

<sup>2</sup>University Technology Malaysia, Malaysia;

<sup>3</sup>Dongguan Power Supply Bureau of Guangdong Power Grid Co., LTD, China;

**Abstract:** With Knowledge management is the transfer of knowledge to people in need. It is the conscious strategy of the organization. Sharing practice-based information in different ways to improve organizational performance. In addition, knowledge sharing is a platform for employees to exchange knowledge and share knowledge within the organization.

**Keywords:** Internal knowledge management; CRM mode

## 1. INTRODUCTION

Successful companies rely more on knowledge management and use knowledge to enhance competitive advantage. However, according to the speed of innovation, companies need to continually update knowledge to create more value for companies, employees and customers. Therefore, learning has become the fundamental guarantee for the survival of an enterprise. The ability of an organization member to acquire knowledge which becomes the core skills of the organization (amine, 2012). In addition, knowledge is the basis for the company to obtain the competitive advantage which is an essential scarce asset of the company.

## 2. INTERNAL OF KM

Knowledge management is the process of obtaining the competitive advantage by identifying, sorting, transferring, and managing knowledge. Including external knowledge and internal knowledge (Gao, Li and Clarke, 2008). External knowledge is related to the knowledge owned by the organization and the knowledge of associated institutions, such as suppliers, users, and other enterprise networks in the same industry (Hall, 2006). The organization's own knowledge is internal knowledge and can be divided into five forms:

Forms <sup>a</sup>	Definition <sup>b</sup>
a) Personal knowledge <sup>c</sup>	Past experience and training. <sup>d</sup>
b) Interpersonal relationship <sup>e</sup>	The network of interpersonal relationships. <sup>f</sup>
c) Database <sup>g</sup>	Standardized knowledge is stored in a structured way. <sup>h</sup>
d) Work-flow and Support System <sup>i</sup>	The required support system to complete the task. <sup>j</sup>
e) Products and Services <sup>k</sup>	Organization's knowledge image. <sup>l</sup>

Practices in the organization avoid the department to avoid duplication in work. In order to achieve effective organizational practice, the correct design of the work, the division of responsibilities will build employees desired specifications (Wickert and Herschel, 2001). It ensures professional tasks and applies to the theory of learning curves. Over time, the work tasks of employees become faster that can produce more products in less time than before. However, specialization may make employees

tired of work and dissatisfied. To be overcome through job rotation and training to enrich skills.

Internal knowledge management is essential to group organizational functions based on the product, work, and location. Functionality refers to the situation where the organization is broken down into departments with different functions (Beijerse, 2000). The organization decide to focus or decentralize function. Furthermore, completion of all functions in each branch office, or issue all orders to the branch office in the headquarters, which is also adopt hybrid systems. It makes the job simple and easy to monitor the task. However, this may lead to a long-term command system, slow decision-making and manager rotation because tend to focus on the areas supervise will lead little understanding of the regions to reduce dependence on external talent.

## 3. COMMERCIAL CONFLICTS AND FACTORS THAT LEAD TO CONFLICT(INTERNAL)

The main causes of workplace conflicts:

Poor management <sup>a</sup>	delayed delivery within departments that share information and materials <sup>b</sup>
Unfair treatment <sup>c</sup>	conflicts between employees and management <sup>d</sup>
Changes in management and organizational structure <sup>e</sup>	Changes occur when management changes and new rules and ways of doing things are introduced. <sup>f</sup>

CRM is mainly an enterprise intelligence system to support marketing, Sales, and Service in the enterprise value chain (Bojanowska, 2017). In addition, customers can fully interact with their chosen path to achieve an increase in satisfaction. From a different point of view, the customer relationship management system distinguishes customers and combines the relationship between management companies and customers to achieve the highest level of satisfaction and at the same time efficiently attract new customers (Wang et al., 2013). Therefore, in the SECI model, belonging to the externalization. Convenience, thus providing a wide range of audience.

## 4. EMPIRICAL STUDY OF FACTORS AFFECTING THE ADOPTION OF TECHNOLOGY.

When adopting a technology, it needs to evaluate whether the technology is cost-effective, able to meet market demands, meet customer needs, and be flexible and easy to use. The company is facing complexity and flexibility issues. The drawback of CRM is also lack of flexibility (Bojanowska, 2017). The reason is that they cannot integrate all departments into a central system which can reduce recurring costs, and the system is difficult to understand (Wang et al., 2013). To lead of conflicts, so it requires a deep understanding of the factors that influence adoption.

Technology adoption measures the degree to which companies use these technologies. It can be assessed by assessing the adoption rate or the degree of adoption. The adoption rate is measured by the amount of time required by a certain number of people (Ribiere and Tuggle, 2010). The degree of adoption depends on the number of technologies used and the number of people using them. However, in developing countries, the slow adoption of technology is mainly due to the lack of appropriate policies, lack of machinery, lack of technical skills, lack of financial support and inadequate communication methods (Carneiro, 2000). Therefore, with the current trend, combined with ERP and CRM knowledge management tools, complimentary use, improve knowledge sharing and system management. In order to enhance competitiveness and social value. In the future development, which will combine the advantages of innovation and integration tools to create the online resource platform that is most suitable for the Bandon Group for remote management and operation.

## **5.CONCLUSION**

The company faces integration issues that constitute the risk of internal conflicts within the organization. A complete system needs to be customized to suit the operation, which should consider the identified factors. The system should also be carefully chosen because of reducing the risk to business operations.

personal knowledge network perspective. *Journal of Knowledge Management*, 16(5), pp.829-844.

[2] Bojanowska, A. (2017). A Nationwide Market Research Project on the Use of CRM Concepts in the SME Sector. *Marketing i Zarządzanie*, 49, pp.113-122.

[3] Carneiro, A. (2000). How does knowledge management influence innovation and competitiveness? *Journal of Knowledge Management*, 4(2), pp.87-98.

[4] Chen, K., Razi, M. and Rienzo, T. (2011). Intrinsic Factors for Continued ERP Learning: A Precursor to Interdisciplinary ERP Curriculum Design. *Decision Sciences Journal of Innovative Education*, 9(2), pp.149-176.

[5] Gao, F., Li, M. and Clarke, S. (2008). Knowledge, management, and knowledge management in business operations. *Journal of Knowledge Management*, 12(2), pp.3-17.

[6] Hall, M. (2006). Knowledge management and the limits of knowledge codification. *Journal of Knowledge Management*, 10(3), pp.117-126.

[7] Ribiere, V. and Tuggle, F. (2010). Fostering innovation with KM 2.0. *VINE*, 40(1), pp.90-101.

[8] Wang, X., Dugan, R. and Sojka, J. (2013). CRM Systems with Social Networking Capabilities: The Value of Incorporating a CRM 2.0 System in Sales/Marketing Education. *Marketing Education Review*, 23(3), pp.241-250.

## **REFERENCES**

- [1] Amine, M. (2012). Knowledge management: a

# Research on visual Design of Cantonese Opera Modeling elements

*Yu Shengnan*

*Guangdong University of Science & Technology, Dongguan 523083, China*

**Abstract:** Cantonese opera is a valuable cultural heritage in Lingnan, China, and it is also attached to internationally. Nowadays, China attaches great importance to the inheritance of excellent culture, and Cantonese opera, as a manifestation of local culture in Lingnan region, plays a self-evident role in the traditional cultural system. At the same time, Cantonese opera also provides rare materials for the development of tourism products in Guangdong. If we extract the visual design creativity from the cantonese opera modeling elements and integrate it with tourism products, it will become a new means to promote the cultural inheritance and promotion of Lingnan region, and also better develop the tourism industry in Lingnan region.

**Keywords:** Guangdong Cantonese Opera; Modeling Element; Visual Design

## 1. INTRODUCTION OF CANTONESE OPERA IN GUANGDONG

Guangdong Opera has always been loved by people. It reflects the charming Lingnan culture, the form is rich, and the cultural connotation is profound. It is popular in the world. The visual design of the elements of Cantonese opera is also a reflection of the unique charm of Cantonese opera. If we can study it and use it to tour product design and development characterized by Lingnan culture, it will become a very bright "element" in tourism products. It is also possible to effectively promote the development of cultural creative industry and tourism.

This article analyzes the visual design of Guangdong Cantonese model elements, some insights are proposed in different respects.

Cantonese opera, it also called the Guangdong Opera, Guangdong Dada, which is the largest drama of Guangdong and Guangxi dialect area. It originated in Foshan and is sung in Cantonese dialect, one of the traditional Operas of the Han nationality. The Cantonese opera is formed in Guangdong, and it has been introduced to Guangxi, Hong Kong, Macau, Taiwan. There are also Cantonese opera performances in China and other countries in Southeast Asia. One of the world's intangible cultural heritage.

Before 1912, the Cantonese opera performance has only been used in Guangzhou dialect. The performance system is getting more and more perfect, and began to wear an exploration of the migrant song in the performance, and change the fake singing as "flat throat" singing. Since 2003, the governments of Guangdong, Hong Kong and Macao have designated the last Sunday of November as Cantonese Opera Day to broaden the audience, promote and preserve Cantonese opera. May 20, 2006, it was

included in the first batch of national intangible cultural heritage list, its number iv - 36. On October 2, 2009, Cantonese opera was listed on UNESCO's Representative List of intangible Cultural Heritage of Humanity.

## II.THE MAIN IDEAS OF VISUAL DESIGN OF CANTONESE OPERA MODELING ELEMENTS

Next, we focus on Guangdong Opera and use the method of refining plane visual elements<sup>[1]</sup>, analyze the main ideas of visual design of its modeling elements.

Currently, Guangdong Cantonese drama is not only loved in a large population, but some young people have also had a strong interest in Guangdong Cantonese opera. We need to understand that if we want the younger generation to better understand and recognize Guangdong Opera, we must innovate the form of Cantonese opera by focusing on their aesthetic taste. In different Cantonese operas, some have more visual beauty, or some have more humorous themes and modeling elements. If we can extract these contents through relevant creative design and make them into exquisite articles for daily use and tourist souvenirs, we can further arouse their appetite for Cantonese opera.

### (I)Choice elements of Cantonese drama characters

When extracting visual elements, we need to follow the guideline of having a clear purpose and system, rather than blindly selecting one or a few characters. Otherwise, it is easy to cause messy, then the resulting product is difficult to form.

In this regard, we will be discussed in terms of two Cantonese opera characters. The first is the famous actor of Cantonese opera -- Red Line girl, as the research object. The second is the "ugly" image with a sense of humor and story, as the creation object. In order to meet the taste of contemporary young people, it is specially designed as a cartoon image.

### (II)Methods and processes for extraction and design

#### 1.Methods of extraction and design

As the "Master" level of Guangdong Cantonese Opera, She can be said to have created the most influential style of singing in the history of Cantonese opera - red school art. This contribution in the history of Cantonese opera is huge.

In the process of creation, we specially selected four representative plays of red School art -- Zhaojun Out of The Gate, Mountain Village, Guan Hanqing and Li Xiangjun, and selected some classic characters from them.

#### 2.Creation process

First, from these classic characters, choose a relatively representative picture. Secondly, observe and determine the specific characteristics of the character image, and outline the general image with a pencil. Again, through the computer to scan the pencil draft, using Coreldraw

software outline and color. Finally, for the parts that could not be trimmed in place before, post processing was carried out through PS software.

The second part is also the most difficult part. Because if we don't have a complete grasp of the content of the story, it is difficult to fully grasp the temperament and charm of the characters. Although it is designed as a cartoon, the connotation of the character must be reflected from it, with a sense of humor and character temperament, rather than just a far-fetched arrangement of a cartoon head.

In the "Mountain country", by the red line girl to play the company commander, image reflects the valiant female soldiers and their great feelings to actively participate in the defense of the motherland. The pose of the character of the character, the fruitful personality has made a vision of vision.

When we design the images of five leading maids in the above classic dramas, we can combine these images effectively, equipped with brief text and concise titles. In this way, a Lingnan local customs as the theme of the postcard design success. Pay attention to the colorful color in the design. The "red" character and the image of "pipa" are integrated, and the "red star" is embedded in the "female" character, in order to reflect the elements of revolution. This not only the title looks more rich, but also make postcards more plot sense, it will to a certain extent to bring a good visual aesthetic feeling<sup>[2]</sup>.

### (III) Extraction and design points of visual elements of ugly characters

Another theme of Cantonese opera is the creation of Ugly Life. Choose three Cantonese opera ugly classic images, the ugly image is humorous and lively, which makes people laugh when they see it. Combining with modern cartoon image processing techniques, it is easier to arouse the interest of the young generation. The creation technique also uses the extraction of typical facial features and detailed description of vivid body posture, combined with pencil draft and graphic software, to create the following three lovely ugly images.

Finally, these three ugly images with their own characteristics were integrated together, with a lively pink as the background color, accompanied by text and title design. Combine into a "ugly" classic image postcard. "Ugly" two word title and caption are using funny shape, with ugly funny feeling.

In addition, in addition to postcards, the author also used the ugly images he created. We want to design two dolls, they are "ugly pillows" and "ugly hanging". The ugly pillow continues the ugly image extracted from the previous article, and summarizes the basic image of the ugly with a few simple large color blocks. The body size of the pillow is moderate, and it will have a humorous effect when young people hold their hands and take photos. It is also comfortable to use it as a cushion or pillow, so

the doll has both practicality and aesthetics.

The creative concept of ugly hanging dolls comes from the cleverness of the hanging function with the "small braid" of the ugly head, achieving the effect of practical and humorous highness. If the doll hangs on the backpack or the bag, it will lead to the eye-catching and created a funny atmosphere.

### III\WHAT SHOULD BE PAID ATTENTION TO IN VISUAL DESIGN

Through the above design methods, the visual elements of Cantonese opera characters are extracted and applied to the design of modern products, so as to achieve the purpose of inheritance and innovation of intangible cultural heritage art. The author summarizes the following design and creation methods:

First, the visual vision element selected must be purposeful and systematic, so that product creation of the later use elements, especially the series of product creations, it will be an orderly and systematic to avoid blind design.

Secondly, when refining and performing a plane effect on the character image, you must first understand the storyline background of the character's role, in order to portray the charm of the characters in place.

Thirdly, in product creation and design, if the functionality can be integrated into the image of the visual element itself, it will effectively enhance the fun of the product.

Finally, the design of the product must comply with the concept of "easy production, reducing cost" and other concepts of industrial production, avoiding high costs such as excessive process techniques. After all, the product is not a handicraft, and it must be a sense of design and convenient production, and the cost can have a wide range of markets and adapt to the level of consumption.

### 4.CONCLUSION

Cantonese opera is the precious cultural heritage of Lingnan region in China. It embodies the charming Lingnan culture with rich forms and profound cultural connotations. It is widely popular all over the world. As for the visual design of elements of Cantonese opera modeling, if we can make full use of it in the design and development of tourism products featuring Lingnan culture, it will become a very bright "element" in tourism products. It is also possible to effectively promote the development of cultural creative industry and tourism.

### REFERENCE

- [1] Yin Xiaowei. Research on plastic arts in traditional cantonese opera costumes [J]. Drama house, 2019, 3 (21): 98-100.
- [2] Xu Xiaoxing. Application of Foshan Cantonese Opera Culture in Network Anime Design [J]. Volkswagen, 2018, 3 (15): 100-101.

# Study On Effective Classroom Training of Machining Specialty

Chen Liang

Guangdong University of Science & Technology, Dongguan, Guangdong 523000, China

**Abstract:** At present, China's economic construction needs many technical practical talents, and higher vocational schools are the main positions of national delivery technology and practical talents, and their teaching directly affects students' professional skills. Machining is a strong course in theoretical and practicality. In the course teaching, in addition to professional knowledge learning, we must pay more attention to practical teaching, and effectively improve the mechanical processing practice of higher vocational students. In this regard, the paper discusses the innovation strategies of practical teaching of machining course in higher vocational colleges based on the shortcomings of practical teaching of machining in current higher vocational schools.

**Key words:** Machining; Training; Effective Classroom

## I. ANALYSIS ON THE SIGNIFICANCE OF TEACHING REFORM OF MACHINING PRACTICE IN TECHNICAL COLLEGES

At present, due to the characteristics of the current stage of development, students majoring in automatic machining and basic machining are required to have higher skills. So, in order to adapt to the needs of talent markets and social development, in the process of teaching development, the mechanical processing major in higher vocational colleges should reform the practical training teaching and construct a more efficient and scientific teaching mode, so as to cultivate talents in line with the needs of social development. Fundamentally improve the employment rate of students majoring in machining.

With the rapid development of our national economy, we have attached special importance to the development of machinery manufacturing industry, which occupies an important position in the current and future development process, for which we have increased financial resources and political support. With the rapid development of industrial modernization, the demand for highly skilled labor in the machining industry has increased not only in terms of quantity but also in terms of professional skills. The necessary technicians can better control advanced production equipment and high levels of skills to achieve high value-added products. However, the current teaching and practical learning of machinery processing in China's technical institutions is not very adapted to the development needs of enterprises. This leads to the shortage of high-quality talents in mechanical processing industry and technical colleges, which is difficult to meet the production needs of enterprises. It has a significant impact on the development of China's

mechanical processing industry and brings serious problems to the employment of talents in colleges and universities. Therefore, this paper briefly analyzes the teaching reform and practical training in mechanical processing in technical schools, in order to cultivate high quality talents in line with the requirements of enterprises.

## II. PROBLEMS EXISTING IN MACHINING TRAINING TEACHING IN HIGHER VOCATIONAL COLLEGES

(I) The teaching content is single and the teaching mode is traditional

The teaching contents and training subjects of machining specialty in higher vocational colleges need to be improved. Traditional teaching content pays attention to the cultivation of student's basic skills and it has strong professionalism, but it is not enough to translate the training skills to training skills. The training content is single, and it is not possible to meet the needs of machinery processing professionals. Although the practical training teaching model is constantly improving and perfecting, but the practical training goal is not clear, it can not stimulate students' interest in learning.

(II) The training ability of teachers needs to be improved, and the overall quality is not high

Most high-skilled personnel are reluctant to teach vocational schools, where people have been outside the year for many years. Another reason is that young teachers go straight from school to vocational schools. They have no skills and experience in the actual training, but also lack the theoretical knowledge that is not conducive to practical teaching. Due to the shortage of teachers, it is especially difficult to provide practical training in higher vocational courses in mechanical processing. Even if it is reluctantly provided, it will not have an effective teaching effect on students. Therefore, vocational schools should pay more attention and take appropriate measures to solve this problem.

## III. PRACTICAL TEACHING INNOVATION STRATEGY OF MACHINING COURSE IN HIGHER VOCATIONAL SCHOOLS

(I) Innovate the content and form of machining practice teaching

In mechanical processing practice teaching, scientific practice content and form are core conditions for ensuring practical effects. Therefore, to change the old and backward practice content and form, and actively improve and innovate is the inevitable way to carry out the machining practice teaching smoothly. First of all, in terms of practical content, higher vocational schools can not limit the content of practical teaching to textbooks, because the content of textbooks is updated slowly and

outdated, which does not meet the needs of students' career development. Higher vocational schools should comply with the times, timely understand the current demand of mechanical processing talents, such as demand, work and technical side. Then adjust the content of practical teaching in time according to the information, formulate scientific practical teaching plan, set up corresponding practical courses, to ensure that students can improve their practical skills in a targeted way, and lay a foundation for future employment and development. Secondly, innovation practice teaching form. In order to change a single problem in the form of practical teaching, teachers can introduce new teaching methods such as multimedia. For example, when explaining mechanical processing knowledge and operational skills, multimedia can be displayed to students through three-dimensional pictures or 3D animations, so that students can make more vivid and intuitive reception knowledge and improve practical teaching efficiency. In addition, teachers can also innovate teaching organization forms, such as introducing hierarchical teaching, group cooperative learning and other modes, so as to enrich practical teaching and enhance students' interest in participating in practice.

#### (II) Pay attention to the improvement of teachers' practical teaching ability

Although the teaching of continuing education is completely entrusted to students, there is no teacher's guidance, but the educational goal cannot be realized, and the teacher is an important existence in the process of continuing education. Therefore, teachers play an important role in practical training teaching, only qualified teachers can guarantee the quality of teaching. Due to the limited number of teachers, vocational education institutions pay special attention to the overall quality of teachers when recruiting teachers, and can only recruit teachers if they meet the required standards in all aspects. With regard to in-service teachers, schools can regularly organize teachers training, and organize research seminars to strengthen and improve teachers' teaching and vocational skills. At the same time, teachers should take the initiative to improve their overall quality and ability. Due to the characteristics of the manufacturing industry, teachers who provide practical training have ideology and requirements that make them lifelong learning. [8] By requiring teachers to constantly develop their own knowledge to meet teaching and learning needs, they can also observe the teaching content and teaching process of excellent teachers to gain experience, learn from each other and cooperate with each other, so as to effectively improve.

#### (III) Expand the students' thinking space and improve their imagination

In order to improve the quality and effectiveness of mechanical transformation of higher vocational practice training, it is necessary to pay attention to cultivating students' imagination and thinking. In order to enlarge thinking space and cultivate students' imagination, teachers can improve teaching efficiency from the perspective of teaching practice and from every aspect of

life. For example, students are encouraged to consider chess trends and the impact of the loser's position on the situation in a variety of ways when playing chess, or to develop sound strategies based on the opponent's situation when participating in competitive sports. Targeted life training can not only help students cultivate good thinking skills, but also stimulate students' interest and enthusiasm to improve the teaching quality of practical training.

#### (IV) Carry out depth school-enterprise cooperation

At present, most of the training and teaching venues and equipment in higher vocational schools are difficult to meet the needs of running schools. Due to various reasons, the training equipment is not updated in time. Effective solutions to this situation are to further strengthen cooperation between school enterprises. On the one hand, when students enter enterprises, they can not only get in touch with the most advanced mechanical equipment in the industry, but also understand the real working situation and requirements of relevant positions in the industry, which can improve their understanding of future jobs. The training masters of the enterprise can combine the implementation of the relevant training knowledge, the history and development prospects of the industry, so that students have established their clear learning goals and stimulate students' learning motivation. The school training teachers can adjust the teaching plan and innovate teaching methods according to the progress of the school-enterprise cooperation. For example, when the training teachers are teaching CNC turning, they can compare the advanced CNC lathes of enterprises and the traditional horizontal lathes of the training room of higher vocational schools. Let students understand the advancement of CNC lathes in the process of craftsmanship, the difference between the normal lathe and CNC lathes and the theoretical and practical knowledge points that operate the CNC lathe. On the other hand, vocational schools can invite enterprise technicians into the classroom, and carry out communication with the school's training teachers. This helps training teachers to improve the training program, improve training teaching methods, so that training teaching is closer to the real situation of the enterprise, helping students' skills, allowing students to adapt to corporate positions faster. Therefore, strengthening school enterprises can effectively improve the quality of training teaching in higher vocational machinery.

### **4.CONCLUSION**

In summary, the improvement of mechanical processing practice in higher vocational schools is significant, and it is the basis and guarantee of students' future employment. However, there are still many problems in practical teaching. Therefore, teachers should attach importance to the innovation of machining practice teaching and improve it from the aspects of practice concept, content, form, platform and teachers. Continuously improve the quality of practice, enhance students' expertise and skills, and deliver high-quality technical practical talents for the state.

**REFERENCE**

- [1] Shaow Xia. On the Problems and Solution Strategy of Teaching of Machining Training in Higher Vocational Machinery [J]. Scientific Advisory, 2019 (044): 27.
- [2] Cheng Lin. Discussion on the Teaching Reform and Practice of Machinery Processing in Higher Vocational Education [J]. Asia Pacific Education, 2019 (10): 130-131.
- [3] Wang Long. Research on Current Situation and Countermeasures of Practical Teaching in Machining Specialty [J]. New Generation: Theory Edition, 2019 (003): 16.
- [4] Liu Yunlong. Sricce exploration of the teaching of mechanical processing training [J]. Scientific Popularity: Science Education, 2019 (6): 141.
- [5] Wu Guangfeng. Exploring how to strengthen the reform of mechanical processing training [J]. Journal of Science and Technology, 2019 (35): 171-172.

# Reconstruction design of environmental facilities in old communities in Dongguan based on Aging Society

Liang Zhu

*Guangdong university of science and technology, Dongguan, Guangdong 532000, China*

**Abstract:** China is now carrying out comprehensive transformation of old communities, and Dongguan is also actively promoting the transformation of environmental facilities in old communities. In the transformation process, many communities pay different attention to the transformation. Some pay more attention to the renewal of community infrastructure and lack of consideration for the living needs of the elderly. Now that China has entered an aging society, it should strengthen more convenience for the elderly to live. The facilities and equipment suitable for the residence of the elderly should be invested more in technology and cost, so as to improve the living community environment of the elderly qualitatively.

**key word:** Old community; The aged; Transformation of environmental facilities

## 1. INTRODUCTION

The 4th Dongguan urban and Rural Planning Committee held its 4th meeting, deliberated and approved the rural construction plan of Dongguan (2018-2035). In the future, 553 villages (communities) in Dongguan will be built into "demonstration area of high-quality urban-rural integration and charming residence in the bay area". The beautiful community action mainly focuses on the needs of customers who have stayed in the old community for more than 5 years, promotes community transformation, improves the quality of the park and improves the lives of residents. Some enterprises are also involved in the transformation of communities, mainly including water seepage and damage of external walls, transformation of fire pipe network, intelligent equipment, maintenance and transformation of ground and wall of underground garage, repair of street lamps, vehicle and pedestrian roads, supplement and repair of return line facilities, green planting and planting in the park, repair of old sports and amusement facilities, etc.

## 2. CURRENT SITUATION OF OLD COMMUNITIES

### 2.1 insufficient barrier free facilities

Restricted by the ideas of the times and building materials, the current community lacks barrier free facilities, which is more obvious for more and more community withdrawals of the elderly. For example, more steps and inconvenient wheelchair access have brought many hidden dangers to the action of the elderly.

### 2.2 the accessibility of public space design in the community is poor.

In terms of appearance and layout, we pursue novel modeling, and let the public landscape and some activity spaces adopt the method of sinking or lifting the ground,

so as to make the overall modeling more novel and beautiful, which makes it difficult for wheelchairs and carts to pass through, which hinders the use frequency of some elderly people.

### 2.3 insufficient types of communication places and lack of aging facilities

There is a general lack of communication places in the residential area, such as public space for the elderly to walk, exercise, rest and communicate. Even if there is such a space, the type is relatively single and lack of diversity. In terms of public facilities, most residential areas have insufficient or unreasonable public facilities.

### 2.4. The landscape design lacks practicability

In order to pursue visual beauty and high greening rate, residential areas generally carry out large-area greening design, usually equipped with small trees, shrubs and green space, but this landscape design will hinder people's sight and affect people's communication. But the elderly have no other choice. They often gather and communicate in such places. At present, most public spaces appear in the form of central square. Although the winding path is very decorative, it is not practical and is not conducive to the passage of wheelchairs for the elderly.

## 3. THE COMMUNITY MEETS THE NEEDS OF THE ELDERLY

Meeting the needs of the elderly refers to their physical and mental health needs, communication and entertainment needs and spiritual and cultural needs. In order to build an aging residential public space, we need to change the conversion conditions of different types of public activity space from these three aspects, so as to lay a solid foundation for the aging transformation of residential public space.

### 3.1 physical and mental health needs

Due to the aging characteristics of the elderly, ensuring the safety of the elderly has become the primary demand of community transformation. The safety of the space outside the residential area is related to whether the elderly have enough confidence and enthusiasm to carry out outdoor activities. In terms of mental health, the elderly have the needs of sense of belonging, neighborhood relationship, sense of security and so on. Most old people need to live in their original homes, because it is essential to be familiar with the environment and care for friends and neighbors. Such a living and communication environment can make the elderly have a sense of belonging and security, and can alleviate the elderly's sense of loneliness and maladjustment, strangeness and loneliness, a sense of loss to the surrounding environment

to a certain extent.

### 3.2 communication and entertainment needs

Neighborhood communication is an important way to establish the sense of belonging and realize self-identity of the elderly. With the change of social roles of the elderly, they often have a sense of self doubt, loss and loneliness. The best way to alleviate these negative emotions is to establish good neighborhood relations and share life experience and cultural cognition. Therefore, the design of communication places in residential areas should pay attention to the communication and entertainment needs of the elderly in public space. In order to attract the elderly to travel, the public space in the residential area should provide rich and colorful leisure activities and space.

### 3.3 spiritual and cultural needs

In addition to meeting the basic needs of physical and mental health, communication and entertainment, the elderly also need to pay attention to the needs of spiritual culture. The spiritual and cultural needs of the elderly mainly include historical and cultural needs, religious and cultural needs and folk cultural needs. These requirements require that when designing the public space of residential area, we should pay attention to giving the cultural attribute and spiritual comfort to the space for the elderly.

## 4. TRANSFORMATION STRATEGY FOR AGING SUITABILITY OF PUBLIC SPACE IN RESIDENTIAL COMMUNITY

### 4.1 safe and barrier free walking environment

The primary problem of community aging transformation is the safety of the elderly, which is the premise and basic requirement of aging transformation. Safety issues should be systematically considered to ensure a safe and barrier free walking environment for the elderly.

#### 4.1.1 transfer of personnel and vehicles in the outer space of the residential area

It is unrealistic to completely avoid the diversion of people and vehicles in the existing community, but the intersection and mixing of pedestrian flow lines and vehicle flow lines should be avoided as far as possible, so as not to affect the internal life of the residential area. For example, carry out traffic control at different times, open parking lots at the entrances and exits of residential areas, re plan traffic flow routes, and try to avoid mixed flow of people and vehicles. In addition, the driving speed of vehicles in residential areas shall be limited, which can be achieved by setting roads, adding green belts, setting speed bumps, etc.

#### 4.1.2 improve accessibility in walking environment

Most of the elderly complete their daily activities by walking, and the scope of daily activities of the elderly is gradually reduced to the community. Therefore, we should pay attention to the community walking system and its supporting barrier free design. First, re plan the walking system of the existing community and increase the consideration of accessibility. In order to green the crowded sidewalk, it shall be widened appropriately to facilitate the use of wheelchairs. Second, carry out step-by-step field design. Ramps are added in the residential entrance, curb, central square and other activity spaces, railings are set on one side, and rest benches are set at

many places along the way to facilitate the passage and rest of the elderly.

### 4.2 construction of public space

The outer space of the residential area provides conditions for the entertainment, communication and fitness of the elderly. We should pay attention to the characteristics of openness and sharing, respect the spontaneous space for the activities of the elderly, give priority to the use needs of the elderly, and take care of other groups at the same time.

#### 4.2.1 enrich public space types

The public space of the existing community is often single, while the outdoor activity space of the elderly includes social space, fitness space, rest space and entertainment space. From the perspective of behavioral psychology, the elderly prefer to integrate these activities. Therefore, these spaces should not have obvious boundaries or too far apart, and the line of sight should be blocked. It shall be arranged adjacent and compatible with multiple functions, such as fitness, chat, chess and card activities, etc.

#### 4.2.2 increase interaction in existing public spaces

Limited by the existing area, some communities cannot add other public spaces. Therefore, it is necessary to use the "gray space" of some communities for certain transformation to increase the communication and sharing of public spaces. For example, add a seat and landscape sketch under a pavilion or shade.

#### 4.3 quiet and comfortable landscape environment

Landscape environment is an indispensable part of the external environment of residential area. Improving the landscape of the residential area can not only improve the microclimate of the residential area, but also provide comfortable rest and viewing conditions for the elderly, which is of great help to increase the possibility of the elderly going out.

#### 4.3.1 increase the connection between the elderly and landscape greening

At present, although most residential areas have a lot of landscape greening, they are relatively closed and lack of hard paved roads. The central landscape is often blocked by a certain line of sight. All these hinder the contact between the elderly and the landscape. Whether it is green plant landscape or water landscape, these are the landscape environmental elements loved by the elderly. Therefore, it is necessary to carry out layered design for the landscape of the existing residential area to enrich the landscape elements and greening types.

#### 4.3.2 pay attention to the enclosure of landscape layout

Closed landscape greening can not only produce more positive space, but also bring a sense of psychological security and sureness to the elderly. Therefore, the site design should form a closed layout as far as possible, and set green or landscape walls behind the seats or at the corner of the site, which can not only improve the privacy of the site, but also provide a hidden rest environment for the elderly.

## 5. CONCLUSION

Our society attaches importance to the care of the elderly, which is the embodiment of the people-oriented spirit. The aging transformation of the external space of the

community can improve the living environment of the elderly, improve the quality of life of the elderly and create a harmonious community atmosphere. This is of great significance to the elderly and the whole society.

#### REFERENCES

- [1] Zhao Wei, Zhao min. from residential area planning to community planning (Journal of urban planning, 2002) China, p 68-71.
- [2] Sooichi Matsumura: Regeneration of residential areas - rebirth of European and American collective housing (Machinery Industry Press, 2008) China, p 50-53.
- [3] Bradford Perkins, J. David hoglen. Residential buildings for the elderly (China Construction Industry Press, 2008). China, p 68-89.
- [4] Jan Gail. Communication and space (China Construction Industry Press, 2016), China, p.50-51.
- [5] Norbert Schutz. Spirit of place (Huazhong University of science and Technology Press, July 2010), China, p.31-41.
- [6] Xiao Jian, Shen decan. Geriatric psychology (China social publishing house, 2009), China, p.21-22.

# Discussion on Life Cycle Management System of Power Grid Equipment Based on Internet of Things

Zhenyuan Zhang

Department of Electronic and Electrical Engineering, Zibo Vocational Institute, Zibo 255000, Shandong, China

**Abstract:** The life cycle management of power grid equipment must integrate cost management, efficiency management and safety management to manage the life of the equipment. Through the scientific and reasonable application of the Internet of Things technology, relying on various sensors to monitor the panoramic state information of the power equipment, and to realize the correlation with the equipment's own attributes (operation status, previous test data, maintenance status, etc.). While judging the equipment status, it also evaluates its service life, realizes functions such as auxiliary decision-making for cycle cost optimization, achieves the life cycle management of power assets, and improves the real-time and accuracy of equipment diagnosis and evaluation, which is useful for production, installation, logistics, and maintenance. Provide a strong basis for many different links.

**Keywords:** Lifetime Management; Power Equipment; Decision-Making For Cycle Cost Optimization.

## 0. INTRODUCTION

Science and technology is the primary productive force and one of the main ways to promote the all-round modernization of society. With the rapid development of science and technology and the emergence of Internet of things technology, the concept of life cycle management is also changing. Electric power enterprises are asset decentralized enterprises, with many asset use departments, a wide range of use locations, complex structure classification and large coverage. At this stage, the fixed assets management of power grid enterprises is still at a very backward level. There are often inconsistencies in accounts, cards and materials. There is a lack of effective control over various management links of assets, such as procurement, operation, capital transfer, maintenance, allocation and scrapping, which brings great difficulties to the management of fixed assets. Under the constraints of reliability and security, how to use high-tech means to build the core competitiveness of enterprises and improve the level of asset management has become an urgent task for power grid enterprises. The main purpose of this study is to solve these difficulties. In order to improve the management level of fixed assets of power grid enterprises, this study uses RFID technology in the Internet of things, combined with advanced technologies such as computer, wireless network communication, intelligent terminal and software system, adheres to the concept of asset life cycle management, and takes the watt hour meter in power assets as the research object to design

a set of watt hour meter life cycle management system based on Internet of things technology. It can realize intelligent supervision in the stages of procurement, warehousing, inventory, operation, maintenance and scrapping, realize the unity of people, accounts, cards and materials, and help power grid enterprises realize the whole life cycle management of equipment from procurement to scrapping; Truly realize the paperless and addressless asset management; Improve work efficiency, reduce management costs, provide accurate reference data for enterprise investment decision-making and rational asset allocation, improve the information management level of power grid enterprises, and promote the construction and development of smart grid. Firstly, this paper analyzes the research status of Internet of things technology and traditional power asset management at home and abroad, points out the shortcomings of power meter asset management, leads to RFID technology, and expounds the research purpose, research methods and innovation of this paper. Secondly, it discusses and studies the current situation and existing problems of electric energy meter asset management system, and puts forward solutions and functional requirements. Thirdly, the key technologies of energy meter life cycle management system based on Internet of things are introduced and analyzed, and the difficulties and solutions of RFID middleware and database synchronization design are emphatically analyzed.

## 1. OVERALL FRAMEWORK DESIGN

(1) Users can use the interconnected information network platform to realize the support of internal and external business of the power grid company through the high integration of data related to various dimensions of power transmission and transformation equipment, and solve the problems of data storage, retrieval, use, mining and confidentiality. This is Implementation platform for power transmission and transformation equipment in the construction of power Internet of Things.

(2) Platform decision-making is to mine the value of data by integrating power production data, operation and maintenance data, and management data to form an application platform for the management of the Internet of Things for power transmission and transformation equipment. The construction of the ubiquitous power Internet of Things for electrical equipment provides a reliable and accurate algorithm model.

(3) Data transmission is through the Internet, communication network and other basic network facilities, used to realize the wide-area data access and transmission

between the perception layer and the platform layer, and provide a safe and fast network channel for the construction of the ubiquitous power Internet of things for power transmission and transformation equipment.

(4) Equipment monitoring is composed of different sensors, controllers, software applications and local communication networks, which connect tens of thousands of equipment and systems such as power production, transmission, and consumption to realize equipment status information and management processes. The collection and aggregation of information such as power grid dispatching data and the local or remote processing of data are the information collection layer for the construction of the ubiquitous power Internet of Things for power transmission and transformation equipment, which can achieve accurate collection of grid data.

## 2. CONSTRUCTION PLAN

Equipment life includes physical life, technical life and economic life. The physical life is determined by the electrical and mechanical properties of the equipment itself; the technical life depends on the design and is determined by human experience; the economic life is based on the equipment's living conditions, input and output, and economic benefits to determine the equipment's retirement life, which is also a cost analysis of the predicted best useful life.

The construction of the equipment's life cycle management system must comprehensively consider various related factors such as physics, technology, and economy.

(1) At the initial stage of equipment purchase, preliminary estimation of equipment parameters, operating environment, and economic benefit indicators should be carried out, and the requirements for equipment should be proposed based on the estimation results.

(2) When the purchased equipment is handed over, it is necessary to check and accept the basic equipment parameters, factory test, installation status, handover test, debugging data, etc., register the assets in the ERP system, and check the basic account, factory, and Data information such as installation and commissioning shall be filed.

(3) After it is put into operation, use the production management system to supervise, monitor, operate and maintain, and evaluate the operation of the equipment to ensure the normal operation of the equipment. At the same time, during the operation of the equipment, the defects, failures and other problems of the equipment are overhauled, modified, etc. Information carries out the construction of a multi-dimensional large database.

(4) Through the establishment of equipment status assessment model, the parameters that have an impact on the safe operation of equipment are extracted for predictive analysis of equipment status, the equipment operation status is evaluated, maintenance or overhaul of abnormal conditions is performed, and the investment funds are associated with the ERP system. By considering the early economic investment and the economic investment in the operation process, the consequences of equipment shutdown are evaluated and analyzed, the

equipment maintenance and the equipment residual value are compared, and the maintenance of the equipment is evaluated and analyzed to determine whether the equipment is overhauled[1]. The reasonableness of the equipment maintenance cost. If the maintenance is carried out, the maintenance cost is included in the asset management system of the equipment's life cycle. Until the equipment exits the system and is scrapped, the equipment's life cycle operation and asset conditions are analyzed, and the equipment service age is obtained. The present value function, the change trend of the equivalent cost of the whole life year with the age of service, and finally the various calculation results are used to calculate the annual maintenance cost expectation and the interruption cost expectation, give the annual operating cost expectation, and combine the failure rate forecast to give the maintenance strategy, Repair project recommendations, estimated economic investment, to achieve the purpose of long equipment life and less maintenance, and also to achieve the purpose of selecting high-quality equipment from high-quality manufacturers, to achieve the optimal use of equipment in the system, and to achieve lean operation and maintenance of equipment in the long run. Intensive project control and asset life-span management are of great significance.

## 3. EXAMPLE DEDUCTION

Take transformer as an example: According to the research and analysis on the operation of the transformer, the life cycle of the transformer can be roughly divided into five stages [2]. The first stage is the manufacturing, transportation, and installation stages. Before the transformer is put into operation, it is a one-time investment with a large amount; the second stage is the first year of operation, mainly for first inspection and routine maintenance; the third stage is operation 2 to 10 years is a stable operation period, with low failure rate, high economy, fewer repairs, and low operation investment; the fourth stage is a period of operation for 10 to 20 years, when the failure rate is rising, and the economy at this stage decreases and investment increases; fifth The stage is after 20 years of operation, and the operation is approaching the economic life. At this time, it is not economical to continue to use, so the transformer is decommissioned and replaced.[3]

From the equipment purchase and delivery, the equipment model, manufacturer, equipment purchase cost, installation and transportation, handover and acceptance, etc. have all been recorded in the ERP system database. The equipment is put into operation. Test information (factory, handover, prevention, routine), etc. will be archived in the production operation management system. At the same time, various tests, maintenance, and repair costs during the operation will also be recorded in the ERP system. When the equipment has defects or failures again, it is evaluated based on the status of the equipment, and the maintenance strategy is determined according to the evaluation results. Generally, the maintenance cost of type A is about 250, 000 yuan, the cost of type B maintenance is 150, 000 yuan, and the cost of type C maintenance is 50, 000 yuan, combined with production Management safety

requirements, consider the impact on production, safety consequences, power outage consequences, maintenance costs, etc. Use the possibility of failure to analyze the safe operation of equipment, and comprehensively analyze whether the equipment is operating based on information such as the residual value of the equipment in the ERP system and the economic indicators of the operating life of the equipment. Repair, and determine the maintenance strategy. When the decision is not worth repairing, start the equipment scrapping process until the scrapping ends, the life cycle management of the transformer has also realized a closed loop, which can be established through the management of the equipment operating status and the life cycle Based on the economic life prediction analysis of equipment use, this also achieves the goal of grid security, the most economical cost, and the highest comprehensive energy efficiency of the equipment during the entire life cycle of the equipment. [4]

#### 4. CONCLUSIONS

Through the construction of a life cycle management system for power grid equipment based on the Internet of Things, information interaction and application interconnection between various software of the production management system can be realized, which improves the data information interaction of equipment, and at the same time uses big data to determine the economic benefits of equipment and improve It improves the lean management level of equipment, promotes the

value-added of power data assets, facilitates the application of digital informatization and the development of big data analysis, realizes the application of data collection, shortens the data interaction time, and reduces the life cycle cost of the equipment, and improves the equipment Comprehensive utilization level.

#### ACKNOWLEDGMENT

This work was supported by the School City Integration Development Plan of Zibo (NO. 2019ZBXC127).

#### REFERENCES

- [1] Yue Xu, Juan Luo. Application of Power Internet of Things Sensor Technology in Online Monitoring of Power Equipment [J]. Digital Technology and Application, 2019, 37(12):36.
- [2] Lin Jia, Ruitao Feng. Internet of Things Technology in Electric Power Application in asset management [J]. Microcomputer application Use, 2013, 29(8): 156-158.
- [3] Yuan Li, Dan Peng. Research on the Application of Internet of Things Technology in Power Equipment Management [J]East China Technology (Academic Edition), 2016, 9(5):188.
- [4] Yong Liao. The application of the Internet of Things technology in the condition maintenance of power equipment [J]. Urban construction Design Theory Research (Electronic Edition), 2016, 27(36): 162-163.

# A Method of Direct Control Design Based on The Least Squares Iterative Algorithm

Jing Wang

Department of Institute of intelligent manufacturing, Zibo Vocational Institute, Zibo 255000, Shandong, China

**Abstract:** This paper presents the least squares based iterative identification algorithms to deal with the output noise influence of the virtual reference feedback turning (VRFT) approach. VRFT approach is based on the measured data of the open-loop unknown plant, and designs the controller directly which makes the closed-loop characteristics fit the reference model objection demand. However, the output noise affects the designed objective and produces the basis to the controller parameters estimate. Though some effective methods have been developed to reduce the effect of the unmeasurable noise, the least squares based iterative identification method is featured by attractive characteristics such as direct using the one set of the measured date, no need to change the structure of the controller and no need to increase the complexity of the controller. The simulation results demonstrate the effectiveness of the proposed algorithm.

**Keywords:** Data-driven design; Virtual reference feedback turning; Least squares method; Model reference control

## 1. INTRODUCTION

In the past several decades, the controller design methods with the plant model techniques have received much research. However, due to the complexity of the industrial processes, the model of the processes is hard to be identified with a reasonable modeling accuracy, not to mention, to use this model to depict the dynamic response of the process. However, if the distinction between the plant model and the process is not too large, it is also boring to design suitable controller which assures the best closed-loop control performance. For this reasons, direct data-driven control methods using merely input and output data of the open loop plant have been developed. Virtual reference feedback turning (VRFT) approach is one of the typical data-driven approaches. This approach is based on the measured data of the open-loop unknown plant, and the controller is designed directly, which makes the closed-loop characteristics satisfy the **reference** model objection demand. VRFT method is originally proposed in [1]~[4], developed in [5], [6] [9]~[12], and applied in [7], [8]. Although VRFT approach also has obtained the promising results, the output noise plays the serious effects of the controller paraments identification, and several papers have considered this problem. In [3], the two-degrees-freedom controller has been designed to take care of the noising affecting the plant, which the complexity of the controller is increased. [5], [10] consider the effect of the noise factor for the linear systems and the nonlinear systems respectively.

In this paper, due to the existing of the output noise, an improvement of VRFT strategy which can reduce the effect of noise and obtain more accurate controller paraments estimate is presented. This paper uses the least squares based iterative identification method in [13] and [14] to solve the output noise estimate problem for VRFT method. [15] and [16] have used the least squares based iterative identification method to the multi-rate systems for the two-input and multi-input systems with the colored noises respectively. The advantage of the proposed algorithm is that it can use only one set of data to estimate the controller paraments without changing the structure of the controller, and increasing the complexity of the controller. The results demonstrated by the simulation will be validated.

The structure of the paper is as follows. In section 2, the basic idea of the VRFT method is described. Section 3 introduces a least squares based iterative algorithm to estimate output noise parameter. It subtracts the effect of the noise estimate from the estimate of the controller parameters, and obtains the perfect controller parameters estimate. Conclusion and remarks are presented in section 4.

## 2. THE VIRTUAL REFERENCE FEEDBACK TURNING FRAMEWORK

The VRFT method approximately solves a model reference problem in discrete time, using the collected openloop I/O data. It designs a controller without resorting to identification of a model of the progress, and for any given desires the response of the closed-loop consisted by  $P(z^{-1})$  and  $C(z^{-1})$  closely to the reference model as depicted in Fig 1.

$$\hat{\theta}(N) = \arg \min J^N(\theta) \quad (1)$$

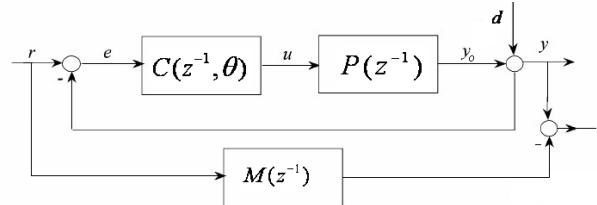


Fig 1. The model reference control system

As stated above, the progress is the unknown plant model, therefore the estimate of the controller parameters vextors is impossible to be obtained. In order to solve this diffiiculty a virtual reference variable is considered.

$$\bar{r}(k) = M^{-1}y(k) \quad (2)$$

where  $y(k)$  is the measured open-loop output data collected form the plant. A virtual is only a virtual variable, using to design the controller, and not using to the actual progress. Then the corresponding tracking error signal can

be obtained. The controller virtual output can be generated by the controller.

It is noted that if given a set of actual measured input signal, a set of output signal also can be collected, even though the actual progress model  $P$  is unknown. A controller can generate  $\{u(k)\}_{k=1, \dots, N}$ , when fed by the error signal. Hence, the idea comes to being. The control objective is changed that searching such a controller  $C(z^{-1}, \theta)$  which fed by the virtual error signal can generate the virtual controller output signal, as closely as the measured actual input signal depicted in Fig 2.

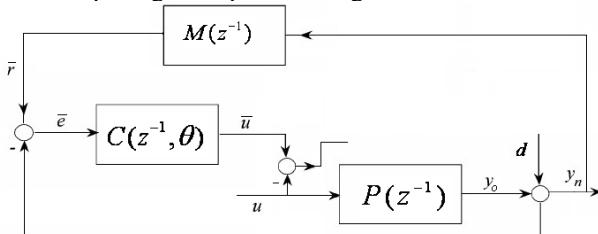


Fig 2. The principle of the VRFT approach

The new control objective is expressed as:

$$\hat{\theta}(N) = \arg \min_{\theta} J_{VRFT}^N(\theta) \quad (3)$$

If the proposed controller is linear in  $\theta$ , then  $J_{VRFT}^N(\theta)$  is quadratic in  $\theta$ . The implementing above idea to identify the controller parameter  $\theta$  is expressed the classical least squares identification. The bulk of the VRFT method is referred to [5].

In the next section, it shows the effect of the output noise  $d(k)$  to the least square estimate, and removes the noise contribution parts from the controller parament estimate by using an iterative approach.

### 3.THE LEAST SQUARES BASED ITERATIVE ALGORITHMS OF EMBEDDED SYSTEMS

The filtered input signal  $u_L(k)$  can be obtained.

$$u_L(k) = L(z^{-1})u(k) \quad (4)$$

Because of exciting of the output noise, the generated virtual reference signal, the filtered virtual tracking error and the information vector  $\varphi(k)$  conclude the noise item. The idea is that the noise contribution parts is removed from the control paraments estimate, and more accurate controller paraments estimation is gotten. Hence, the output data  $y(k)$  concluded the noise  $d(k)$  is replayed with the estimate of the noise-free output  $y_0(k)$ .

The following summarizes the steps of the least squares based iterative VRFT design of controller:

- 1) Collect the I/O data and give the reference model  $M(z^{-1})$
- 2) Generate  $u_L(k)$
- 3) Let  $q = 1$ , initialize the noise and obtain the initialized the noise -free outut
- 4) Form the iteration estimate variable
- 5) Update the iteration
- 6) Compute the iteration estimate variable
- 7) Compare  $q$  with  $Q$ . If it is equal, then terminate the procedure and obtain the  $q$ th iteration estimate, otherwise set  $k = k + 1$ , and go to step 4.

The flowchart of computing the controller parameter estimate in the least squares based iterative algorithm is shown in Fig. 3 with the increasing of the tag  $k$ .

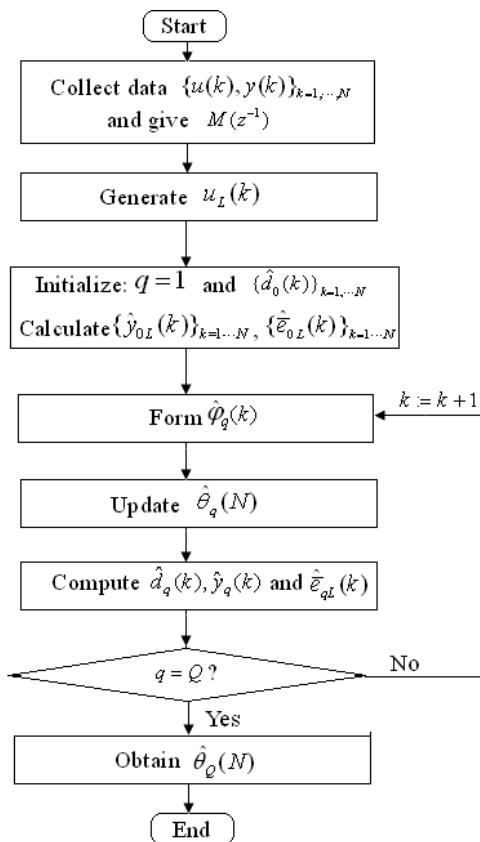


Fig 3. The model reference control system

### 4.CONCLUSIONS

In this paper, we have proposed least squares based iterative identification algorithms to deal with the output noise influence of the virtual reference feedback turning approach. The basic idea is that the noise-contribution parts is removed from the controller parament estimate, and the quite good estimation accuracy is obtained. The advantages of the least squares based iterative identification is that it uses the only one set of the measured date, does not need to change the structure of the controller, and not need to increase the complexity of the controller. The simulation results demonstrate the effectiveness of the proposed algorithm. The results show that an more accurate estimate can be achieved by this approach.

### ACKNOWLEDGMENT

This work was supported by the School City Integration Development Plan of Zibo (NO. 2019ZBXC127).

### REFERENCES

- [1] S. M. Savaresi, and G. O. Guardabassi. Approximate feedback linearization of discrete time nonlinear systems using virtual input direct design, *Systems and Control Letters*, vol. 23, pp. 63-67, 1997.
- [2] H. Nijmeijer and S. M. Savarcsi. On approximat model-reference control of siso discrete-time nonlinear systems", *Automatica*, vol. 34, pp. 1261-1266, 1998.
- [3] G. O. Guardabassi, and S. M. Savaresi. Virtual reference direct design method: An Off-Line Approach to Data-Based Control System Design", *IEEE Trans. Automat. Contr.*, vol. 45, pp. 954-959, 2000.
- [4] M. Stefanovic, and M. G. Safonov. Safe Adaptive

- Switching Control Stability and Convergence”, IEEE Trans. Automat. Contr., vol. 53, pp. 2012-2021, 2008.
- [5] M. C. Campi, A. Lecchini, and S. M. Sabaresi. Virtual reference feedback tuning: A direct method for the design of feedback controllers”, Automatica, vol. 38, pp. 1337-1346, 2002.
- [6] J. Liu, and X. G. Liu. least-squares based iterative parameters estimation for two-input multirate sampled-data systems”, In 2009 American control conference, June 10-12, St. Louis, USA. pp. 4379- 4383
- [7] X. G. Liu, and J. Liu. least-squares based iterative parameters estimation for a class of multirate system”, Automatica, vol. 46, pp. 549-554, 2010.
- [8] Fabio Prebidi, Thomas Schauer, and Sergio M. Savaresi, and Ken J. Hunt. Data-Driven control Design for Neuroprostheses: A Virtual Reference Feedback Tuning (VRFT) Approach”, IEEE Trans. Automat. Contr., vol. 12, pp. 176-182, 2004.
- [9] F. Ding. System identification theory and method and Matlab simulation”, Beijing: China Electric Power Press(in Chinese), 2010
- [10] M. C. Campi, and S. M. Sabaresi. Direct Nonlinear Control Design: The Virtual Reference Feedback Tuning (VRFT) Approach”, IEEE Trans. Automat. Contr., vol.51, pp.14-27, 2006.
- [11] Antonio Sala. Integrating virtual reference feedback tuning into a unified closed-loop identification framework”, Automatica vol. 43, pp.178-183, 2007.
- [12] Yasuki Kansha, Yoshihiro Hashimoto, and Min-Sen Chiu. New results on VRFT design of PID controller”, CHemical Engingeering Research And DesignProc. of American, vol. 86, pp. 925-931, 2008.
- [13] F. Ding, and T. Chen. Identificaiton of Hammerstein nonlinear ARMAX systems”, Automatica, vol. 41, no. 9, pp.1479-1489, 2005.

# Prediction of Fungi's Decomposition Rate Based on Markov Chain and Gaussian Distribution

Hao Ouyang<sup>1</sup>, Zi-Nan Li<sup>1</sup>, Jia-Ming Zhu<sup>2\*</sup>

<sup>1</sup>School of Machine and Engineering, Wuhan University, Wuhan 437000, China;

<sup>2</sup>School of Statistics and Applied Mathematics, Anhui University of Finance and Economics, Bengbu 233030, China

\*Corresponding Author.

**Abstract:** The decomposition of organic material is a critical component of the carbon cycle, and fungi are main contributors in decomposition. Therefore, study the decomposition rate of fungi is important to help accelerate carbon cycle and help protect environment. In this article, we propose a model to predict the decomposition rate of different combinations of fungi in various environments. Firstly, we use k-means method to classify fungi, and build a Markov Chain based model to describe the interspecies interaction; Secondly, we use multiple gaussian distribution to analyze the impact of environment on fungi's decomposition rate. Finally, we combined above two models to evaluate the decomposition rate of different combinations of fungi in various environment. Through Matlab simulations, the model can be applied to predict fungi's decomposition rate and find the most suitable combination of fungi for environmental protection. Additionally, it can verify the importance of biodiversity.

**Keywords:** Decomposition rate of fungi; Markov chain; K-Means; Multiple Gaussian Distribution; MATLAB

## 1. INTRODUCTION

### 1.1 DECOMPOSITION OF FUNGI AND ITS INFLUENCING FACTORS

Fungi are decomposers in nature. They contribute to accelerating the earth's carbon cycle by decomposing woody fibers and ground litter. Therefore, increasing their decomposition rate is of positive significance to the earth's carbon cycle.[1] As the carbon cycle is getting more and more attention, studying the decomposition rate of fungi has become increasingly important. Recently, scholars have studied the main factors affecting the decomposition rate of fungi. There are two factors that affects fungi's decomposition rate according to Nicky Lustenhouwer recent research.[2]

①Internal Factors--Competitive Ranking: There are mainly four types of interspecies relationship between fungi, predation, competition, parasitism and mutualism. If fungi's living conditions are highly overlapped, there will be competition between fungi.[3] Competitive Ranking reflects the degree of competition between fungi's populations. In the case of limited resources, fungi with strong competitiveness can occupy the niche of fungi with weak competitiveness and obtain more resources, while fungi with weak competitiveness is in a suppressed state, which is manifested as a decrease in the

decomposition rate.

②External Factors--Environmental Factor: Fungi's decomposition rate is affected by temperature and moisture.[4, 5] Each fungus has its suitable temperature and moisture zones. Under suitable temperature and moisture, the decomposition rate reaches the maximum. The moisture niche width reflects the degree of fungi' dependence on water, which is reflected in the adapt ability of fungi to the environment.

### 1.2 MODEL ASSUMPTIONS

In order to facilitate the solution of the problem, the following conditional assumptions are proposed: ① Competition is the only interspecies relationship between fungi; ②Competition does not lead to the extinction of fungi, but only leads to inhibition of fungal life activities; ③The decomposition rate of fungi is only affected by competition, temperature and moisture.

## 2. INTERSPECIES INTERACTION IMPACT ON FUNGI'S DECOMPOSITION RATE BASED ON MARKOV CHAIN

### 2.1 FUNGI INITIALIZATION

In order to simplify the model, we selected the most representative four types of fungi for simulation through k-means clustering, and gave the possible distribution densities of various groups. Then our model describes the interaction between fungi based on the Markov chain model.

Table 1. Fungi Type

Type	Growth Rate/Decomposition Rate	Competitive Ranking	Moisture Niche Width
a	High	High	Low
b	High	Low	Low
c	Low	High	High
d	Low	Low	High

Our model uses decomposition rate and moisture tolerance to describe the inherent properties of fungi. Moisture tolerance is the difference between a fungus' competitive ranking and its moisture niche width. According to the research of Nicky Lustenhouwer, fungi with slow growth rate are often more adaptable to changing environment than fungi with fast growth rate. Meanwhile, the growth rate and decomposition rate are positively related and one-to-one correspondence. Therefore, we divided all fungi into four categories according to their decomposition rate, competitive ranking and moisture niche width. Table 1 shows our

classification.

In order to determine the specific scope of each index, we analyzed Nicky Lustenhouwer's research. We found that moisture tolerance ranges from -0.7800 to 0.9900. Define  $W_H, W_L, W_M$  as lower, middle and upper bound of moisture niche width, then low width  $\in (W_L, W_M)$ , high width  $\in (W_M, W_H)$ ; Similarly, low ranking  $\in (R_L, R_M)$ , and high ranking  $\in (R_M, R_H)$ . In accordance with our classification rules, we equally divide the moisture tolerance into four parts -0.7800~0.3375, -0.3375~0.1050, 0.1050~0.5475, 0.5475~0.9900 and get the following equation set.

$$\begin{cases} R_L - W_H = -0.7800, \\ R_H - W_L = 0.9900, \\ R_L - W_M = -0.3375, \\ R_M - W_L = 0.5475, \\ R_M - W_H = -0.3375, \\ R_H - W_M = 0.5475. \end{cases} \quad (1)$$

Both moisture niche width and competitive ranking belong to  $(0, 1)$ .

Multiple and low order equation group do not necessarily have solutions Solving with Matlab, we found that to enable the equation has a solution,  $W_H$  is required to belong to  $(0.885000, 0.895000)$ . Without loss of generality, we set  $W_H = 0.890000$ , then we can get the range of competitive ranking and moisture niche width. To get the exact range of decomposition rate, we need to use clustering which we'll go into details in the next section.

The results are shown in Table 2.

Table 2. Scopes of Each Index

Index	Decomposition Rate	Competitive Ranking	Moisture Width	Niche
high	0.067340	~	0.552500	~
	0.436984		0.995000	0.890000
low	0.065783	~	0.110000	~
	0.386865		0.552500	0.447500

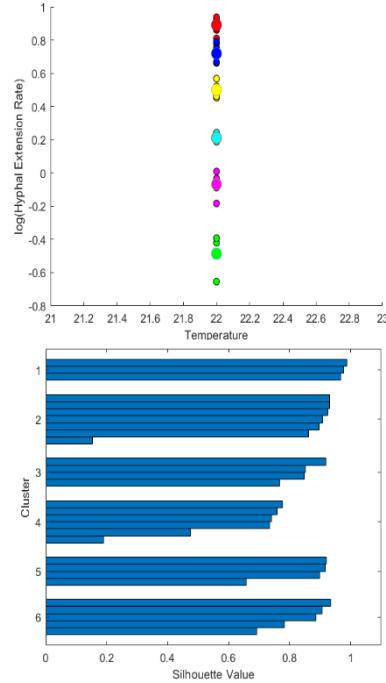
## 2.2 THE K-MEANS METHOD

The k-means algorithm is a method of constructing  $k$  divided clusters according to a given data set of  $n$  data objects, and each divided cluster is a cluster. This method divides the data into  $n$  clusters, each cluster has at least one data object, and each data object must belong to and can only belong to one cluster. At the same time, the similarity of data objects in the same cluster is high, and the similarity of data objects in different clusters is small. Cluster similarity is calculated using the mean value of objects in each cluster.[6]

Besides, k-means seeks to minimize the average squared distance between points in the same cluster, so we can get the centroid of a cluster.[7]

Since the moisture niche of different types of fungi are highly overlapped, which all live by decomposing woody fibers and ground litter, so we only consider the competition between fungi. Competition among populations often has three results: dominant, disadvantaged, and inhibited, so there are two fungi types and three states. In order to get the decomposition rate of fungi, we need to quantify the growth rate range of both fungi first. We analyze the growth rate data at  $22^{\circ}\text{C}$  (suitable environment), because the data distribution is asymmetric, in order to make the data evenly distributed to improve the accuracy, we take the logarithm of the

growth rate, and then use the k-means method to divide it into 6 categories. The clustering results are shown in Figure 1.



(a) Fungi are Classified into Six Groups (b) Clustering Evaluation

Figure 1. Clustering Results

Through clustering, we can divide fungi into six categories according to their growth rate, and we can get the decomposition rate of fungi based on their growth rate. We sort the fungi by decomposition rate from high to low, and the first, third, and fifth groups are used as the three states of the high decomposition rate fungi. Similarly, the second, fourth, and sixth groups are used for the low type. Thus, the decomposition rate range of two types of fungi can be obtained. For each status of both fungi, k-means clustering is used again to obtain the center point of the decomposition rate of the type of fungi, that is, the most likely decomposition rate of the type of fungi. The clustering results are organized in Table 3.

Table 3. Status and Decomposition Rate

Status	High Decomposition Rate	Low Decomposition Rate
Dominant	0.23900	0.17150
Disadvantaged	0.16600	0.11000
Inhibited	0.09954	0.09398

Considering that the decomposition rate of fungi in a certain state is not invariable, we add a random number to the decomposition rate of each type of fungi. Random numbers obey normal distribution  $N(0, 0.01)$ .

## 2.3 THE MARKOV CHAIN MODEL

A Markov chain is a stochastic model describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event.[8] Proportion of three fungal states in population is also only related to the current state. The most important components of Markov chain prediction are the initial state matrix and the transition matrix. The next step state matrix is multiplied by the state matrix and

the transition matrix, and repeats until the state matrix becomes stable.

The Markov absorption chain refers to a factor that can only remain unchanged once generated in the state space, and this Markov chain will converge to the invariant factor.[9] Interspecies competition also exists the condition that a species is inhibited by other species, and the population cannot be restored. Therefore, we can use Markov chain to predict interspecies competition.

Step1: Get the Initial State Matrix of the Markov chain.

In the initial state, the proportion and distribution of the four fungi are random. The proportion of the initial population density will significantly affect the distribution of the initial competitive state.[10] To describe this relationship, let the proportion of the initial population of the four fungi be a: b: c: d (a, b, c, d here means population density of a, b, c, d).

Inspired by cellular automata, [11] we construct a rectangle with three rows and three columns. Focus on the fungi in the center, its initial state is affected by the surrounding eight fungi. As direct competition and indirect competition has different impact on the fungus' state, so we set that the influence of direct contact with fungi is twice that of indirect contact. After normalization, the scores are 1/6 and 1/12 respectively. Surrounding fungi may be more competitive, less competitive or as competitive as the same. We rule that for two fungi with the same competitive ability, the scores are equally divided; for two fungi with different competitive ability, because the average ranking values of strong and weak competitive ability are 0.75 and 0.25, respectively, the scores are assigned at 3:1; for the same kind of fungi, it gets all scores, because we do not consider intraspecies competition.

$$\begin{cases} Scores_a = a + \frac{1}{2}b + \frac{3}{4}c + \frac{3}{4}d, \\ Scores_b = \frac{1}{2}a + b + \frac{3}{4}c + \frac{3}{4}d, \\ Scores_c = \frac{1}{4}a + \frac{1}{4}b + c + \frac{1}{2}d, \\ Scores_d = \frac{1}{4}a + \frac{1}{4}b + \frac{1}{2}c + d. \end{cases} \quad (2)$$

Step2: Get the Transition Matrix of the Markov chain.

We calculate fungi's decomposition rate at 22 degrees Celsius, and calculate the number of each fungi in the three status to obtain the ratio of the absorption probability of each fungus at 122 days. Based on the default conditions(Four kinds of fungi exist, and the ratio is 1:1:1:1), without loss of generality, we set the probability ratio of the dominant and disadvantaged to inhibited as 2:1. Then we can calculate absorption probabilities and transition matrix of each fungi when the initial proportions of the four fungi are different.

Nicky Lustenhouwer's research shows that high competitive ranking fungi will not be inhibited, while low competitive ranking fungi may be inhibited. Fungus with low competitive ranking and low moisture niche width have a higher probability of absorption, and our model reflects this feature. Probability of absorption can be calculated by equation(3).

$$\begin{cases} P_a = 0, \\ P_b = \frac{1}{2}Scores_b + \frac{3}{4}(1 - Scores_b), \\ P_c = 0, \\ P_d = \frac{2}{9}Scores_d + \frac{4}{9}(1 - Scores_d). \end{cases} \quad (3)$$

When we mention interspecies competition, Lokta-Volterra model is widely used to solve such problems, [12] we refer to its two common competition results. For the competition between biological species, the ratio of the dominant and disadvantaged tends to be exchanged every time they go through a cycle of competition. At the same time, species with low competitive ranking may enter a state of inhibition and cannot recover. In the Markov chain, they enter the state of absorption. If the competition cycle changes, the state transition matrix will change. We consider that the competition cycle of fungi is one day. Therefore, the state transition matrix for fungi with high competitive ranking and fungi with low competitive ranking are:

$$M_h = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, M_l = \begin{pmatrix} 0 & 1-x & 0 \\ 1-2x & 0 & 2x \\ 0 & 0 & 1 \end{pmatrix} \quad (4)$$

For the Markov absorption chain, after multi-step conversion, the absorption probability is not sensitive to the initial state matrix, and can be considered only related to the transition matrix, the formula is as follows:

$$p = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} = \begin{pmatrix} 0 & 1-x & x \\ 1-2x & 0 & 2x \\ 0 & 0 & 1 \end{pmatrix}^{122} \quad (5)$$

Using the classification method based on Table 3, it can be seen that the inhibited probabilities after 122 days are 5/11, 1/3 for two types of fungus (b and d, respectively). Now we know the initial state matrix and absorption state probability of the Markov, so we can calculate the transition matrix. But matrix inversion is cumbersome, so we each time let the denominator of transition matrix plus one, and calculate the corresponding the absorption state probability. At last, we record all the data and make a data table. We only use data with absorption probability in the range of (0.25 ~ 0.85), and use function to fit them. The fitting results are as Figure 2.

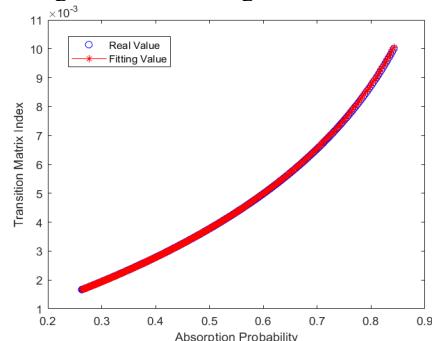


Figure 2. The Relationship Between Absorption Probability and Transition Matrix.

Matrix Index is as follows:

$$x = -\frac{1}{183.6} \ln(1-p) \quad (6)$$

To sum up, we have established the following model: Get the Markov initial state matrix from the Markov initial population distribution, and then get the transition matrix.

## 2.4 PREDICT FUNGI'S POPULATION DENSITY

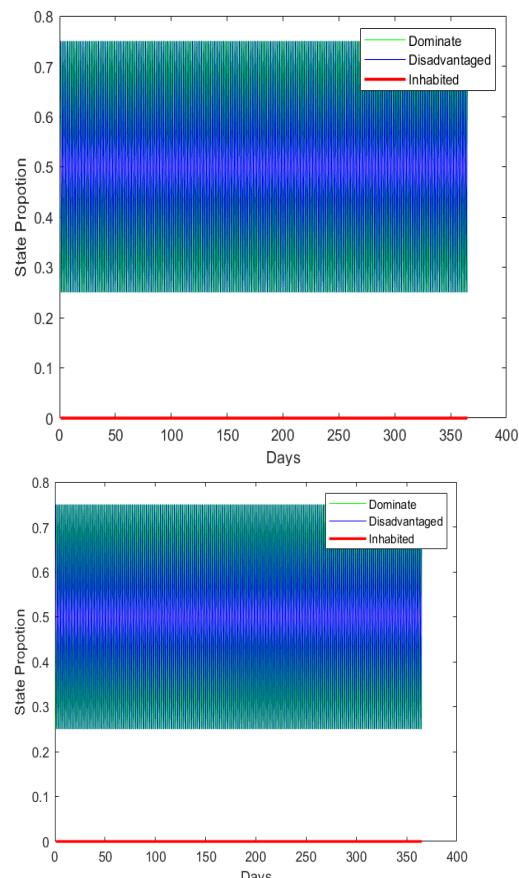
Take the default state as an example, the initial state matrix(show the dominant proportion only) and one-step transition matrix(show the absorption probability only) can be known using equation(6), so we get the Table 4.

Table 4. Initial State Matrix and One-Step Transition Matrix

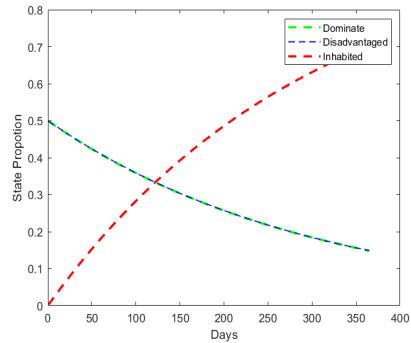
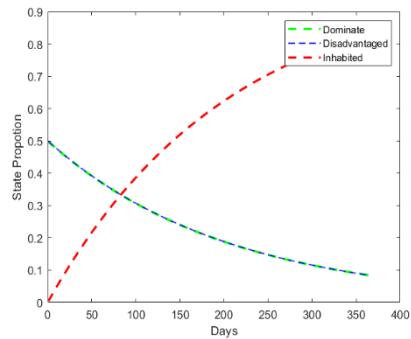
Fungi Type	a	b	c	d
Initial State Matrix	0.7500	0.5000	0.7500	0.5000
Transition Matrix	0.0000	0.0033	0.0000	0.0022

We take 365 competition cycles as an example, that is, to predict the interaction of fungi in a year, the proportion curves of each of the four fungi are shown in Figure 3. Note that the curves in Figure 3(a) and (b) are the same, because the fungi will not be inhibited. In Figure 3(c) and (d), the blue curve and the green curve show a trade-off relationship, which means the fungi are competing fiercely and their competitive ranking are similar.

Due to the mathematical properties of the Markov absorption chain, the fungi are considered unable to recover after entering the inhibited status. In this status, the fungi will lose part of the environmental resources. Considering the different adaptability of the four fungi to the environment, adaptable fungi will expand their ecological width and maintain more resources, so we simply rule that the allocation ratio is 0.60 and 0.40, respectively. As the same, unadaptable fungi will lose part of their resources to adaptable fungi, so we set ratio of lose to be 0.75 and 0.80.



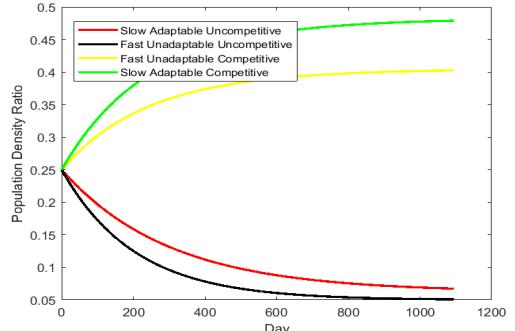
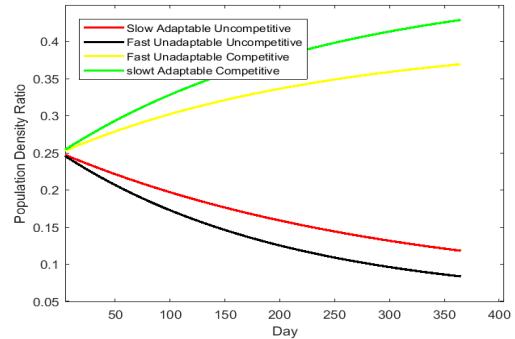
(a) Fast Adaptable Competitive Fungi (b) Slow Unadaptable Competitive Fungi



(c) Fast Adaptable Uncompetitive Fungi (d) Slow Unadaptable Uncompetitive Fungi

Figure 3. Proportion Curves

According to the above regulations, we can predict the population density of the four fungi after one year and three years respectively, as shown in the Figure 4(a) and (b).



(a) Prediction of One Year (b) Prediction of Three Years

Figure 4. Population Density Ratio Changes Over Days  
From the above prediction, it is not difficult to see that the short-term trend of the interaction between fungi is that the proportion of low competitive fungi in the inhibited status is increasing, and the population density is gradually decreasing; the status of competitive fungi is changing between dominant and disadvantaged.

Relatively stable, and the population density is gradually increasing. In the long term, if the environment does not change, fungi with high competitive ranking will occupy most of the ecological niche, and fungi with low competitive ranking will be strongly inhibited, and the population density ratio will be stable.

### 3. ENVIRONMENT IMPACT ON FUNGI'S DECOMPOSITION RATE BASED ON GAUSSIAN DISTRIBUTION

#### 3.1 ENVIRONMENT INITIALIZATION

Moisture and temperature are the most important indices in abiotic environment that affect the decomposition of fungi, so we only take temperature and moisture as the characteristics of the environment and then initialize them.

##### ◆ Moisture Index

Our model uses precipitation to describe moisture. In Table 5, we list five kinds of area type and their corresponding precipitation. (SI: mm/year)

Table 5. Precipitation and Area Type

Type	Arid	Semi-arid	Semi-humid	Humid
Precipitation	<200	200~500	500~800	≥800

The optimal precipitation for fungi is set to 1000mm, [13] as humidity is the most suitable condition for fungi's decomposition. Moisture Index and precipitation are linear correspond. For example, precipitation 1000mm corresponds to 0.5, precipitation 250mm corresponds to 0, precipitation 1750mm corresponds to 1.

##### ◆ Temperature Index

The fungus has an optimal temperature range for decomposition, beyond this range, the fungus will stop decomposing, because they cannot grow in that temperature.[14] Rule that the upper temperature  $T_h$ , lower temperature  $T_l$ , given temperature is  $T$ , [15] then temperature index X is known by equation(7).

$$X = \frac{T - T_l}{T_h - T_l} \quad (7)$$

#### 3.2 GAUSSIAN DISTRIBUTION

According to Shelford's law of tolerance and Liebig's law of the minimum, [16] influence of environment on the decomposition rate of organisms is generally distributed normally or can be approximated by Gaussian Distribution(normal distribution).[17] Based on this idea, we proposed a multiple Gaussian distribution model. The expression for the normal distribution is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (8)$$

where  $\sigma$  and  $\mu$  is the variance and mathematical expectation of the normal distribution.

Once  $\sigma$  and  $\mu$  is obtained, the Gaussian curve can be determined, then the decomposition rate is known.

Step1: Get Gaussian Distribution of a Certain Type of Fungus.

To simulate a certain kind of fungi, we only need to know its suitable decompose temperature and moisture niche width.  $\mu$  corresponds to the optimum moisture for fungi's decompose.

since we scaled moisture index to [0, 1],  $\mu$  is most likely around 0.5, because fungi can maintain as good decomposition rate in both arid and humid area. [18]  $\sigma$

can be derived from the moisture niche width w.

$$\sigma = w * 2^{\frac{1}{2}} \log(2)^{\frac{1}{2}} \quad (9)$$

The specific data are determined by fungi kind.

Step2: Get Gaussian Distribution of Various Type of Fungi. As our model simulate a type of population rather than a specific population, the optimum moisture index and accurate water niche width is unknown to us. We assume that the uncertainty of the width of the moisture niche also obeys the Gaussian distribution. Due to interspecies differences, the most suitable moisture condition for fungi cover all moisture index, so we assume that in  $\mu$  obeys Gaussian distribution of  $N(0.5, 0.5/3)$ , which means that the most suitable fungus' decompose moisture index varies from 0~1 in 99.73% cases.

From Table 3, we can get the range of water niche width. When the population is large enough, the moisture index of fungi obeys the normal distribution. Take unadaptable fungi for example, its decomposition rate ranges from 0.067340 to 0.4369840, so its mathematical expectation is  $(0.436984 + 0.067340)/2$ , and the variance is  $(0.436984 - 0.067340)/3$ . Thereafter, the niche width of unadaptable fungi can be obtained from equation (10).

$$w_{unadaptable} \sim N(0.024875, 0.08290) \quad (10)$$

In order to simulate huge populations, we generate about 10000 curves, and calculate their average values under the certain moisture. By changing the value of moisture index, we can get the corresponding decomposition coefficient under different moisture. In Figure 5(a) we show the relationship between precipitation and decomposition coefficient.

#### 3.3 EXPERIENCE-BASED EQUATION

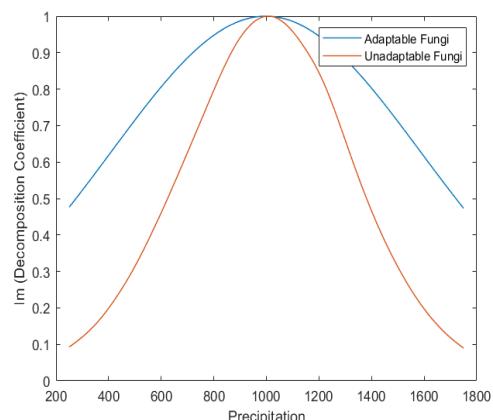
Temperature's impact on fungi's decomposition rate is dissymmetric.[19] From K Chen's research, we know that temperature impact fungus decomposition rate by the following formula:

$$I_t = \sin^2(1.31X - 0.0084X^2 + 0.000133X^3) \quad (11)$$

It here stands for decomposition coefficient by Temperature.

The temperature range for fungi's decomposition is set to be 0~35 so that they can grow. In Figure 5(b) we show the relationship between temperature and decomposition coefficient. We can see that, the suitable temperature for fungi's decomposition is around 23°C.

#### 3.4 ANALYZE ENVIRONMENT IMPACT ON FUNGI'S DECOMPOSITION



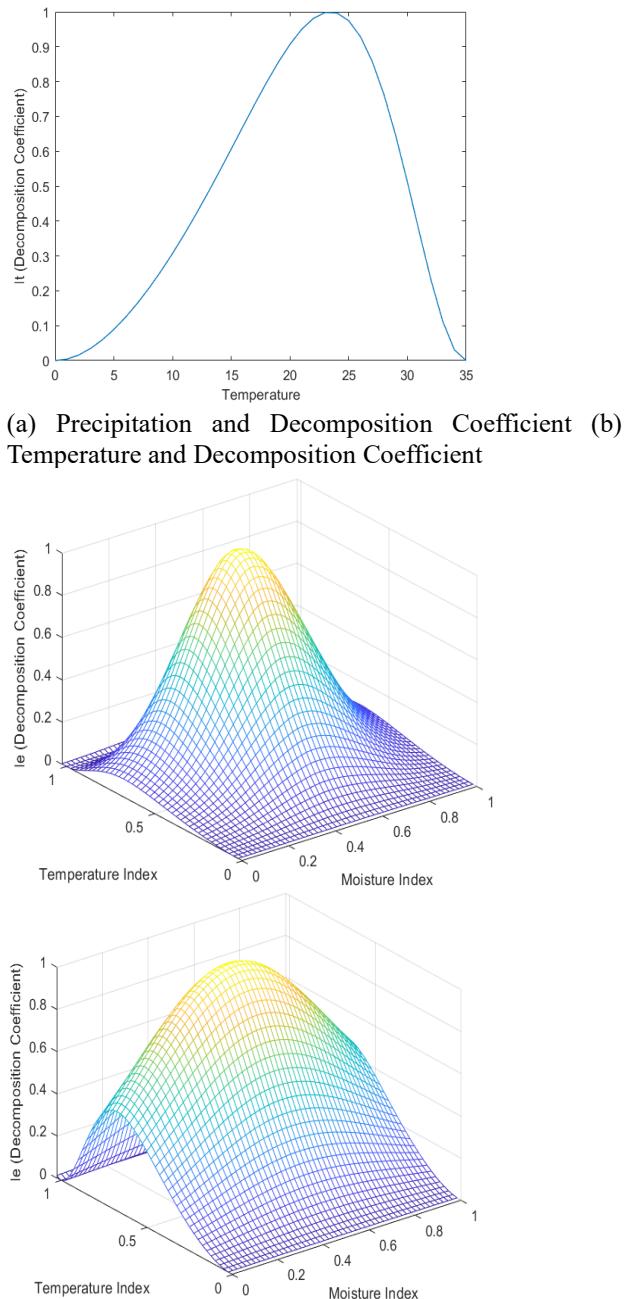


Figure 5. Environmental Index Changes Due to Moisture and Temperature

In order to quantify the impact of environment on the decomposition of fungi, we simplify the combined effects of moisture and temperature, and multiply their effects as environmental decomposition coefficient. The formula is as follows.

$$I_e = I_t I_m, I_e \in (0, 1) \quad (12)$$

We can get Figure 5(a, b, c, d).

Figure 5(c) and (d) presents the environment impact on unadaptable fungi and adaptable fungi separately. Compare to Figure 5(d), the curve in Figure 5(c) is slim and narrow, which means that a slight environment change will greatly affect fungi's decomposition rate.

#### 4. RESULTS AND DISCUSSION

##### 4.1 ANALYZE FUNGI'S DECOMPOSITION

##### COMPREHENSIVELY

In the previous section, we calculate the decomposition rate of the fungus affected by competition. In the competition model, we divide fungi into four types, each of which has three states. So, the decomposition rate of a certain kind of fungus can be described below. Let  $I_c$  stands for the decomposition rate of a group of fungi, and the three values of  $i$  correspond to the three states of dominate, disadvantage and inhabitant respectively. We use Ratio to present proportion of fungi in a certain state in the population, then we can get  $I_c$  through the following equation:

$$I_c = \sum_{i=1}^3 \text{Ratio}(i) \times \text{Status}(i) \quad (13)$$

Considering that the environment also has impact on fungi's decomposition rate, we need multiple  $I_c$  and decomposition coefficient to get the final decomposition rate expression of a certain type of fungus. Environmental factor is defined in equation (12).

Accumulate the all fungus's decomposition rate, we can get the total decomposition rate of the fungus combination.

$$N = \sum_{j=1}^4 I_c(j) I_e(j) \quad (14)$$

The four values of  $j$  correspond to the four types of fungi. In order to reflect the accuracy of the model, we simulate the decomposition of fungi according to the experimental conditions in Nicky Lustenhouwer's research. Under different experimental conditions, final result is similar to the real data, which proves that our model has good performance and robustness. With this foundation, we can then verify the decomposition trend of fungi in various environments and the pros and cons of various species combinations.

##### 4.2 THE DECOMPOSITION RATE OF DIFFERENT COMBINATIONS OF FUNGI IN VARIOUS ENVIRONMENT

Using our model, we can predict the decomposition rate of different combinations of fungi in various environments. We do not simulate all kinds of environment, to simplify the process and make our results more representative, we simulate two kinds of environment. One is suitable for fungi's decomposition, while the other is unsuitable.

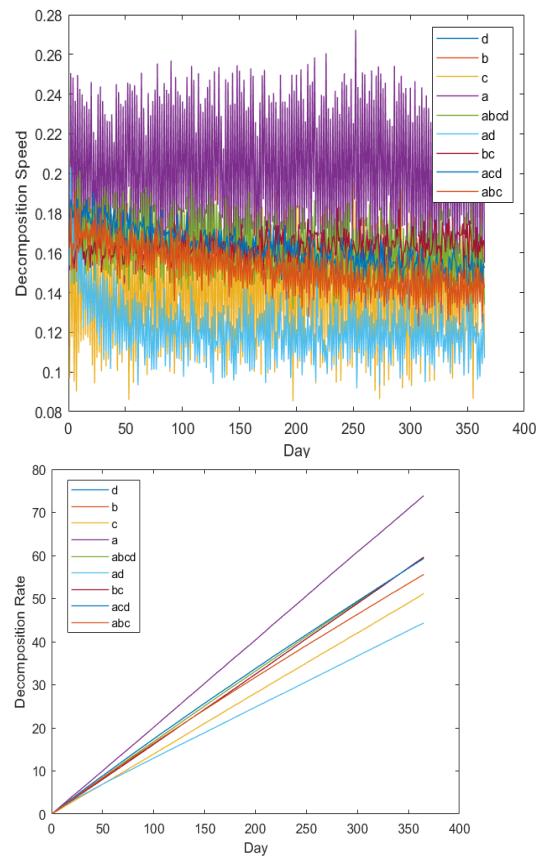
◆ Unsuitable Environment(Arid):[20] The temperature is basically maintained at 25 to 28 degrees Celsius, annual precipitation is generally less than 200mm.

◆ Suitable Environment(Arboreal):[21] Temperate is around 23 degrees Celsius and precipitation is 1000mm. There are four situations when the four fungi are cultivated separately, there are 6 situations in two combinations, 4 situations in groups of three, and one situation in four kinds of fungi, so we have 15 combinations in total. We will exclude six combinations which will not be considered for their characteristic can be predicted. For example, a and b's decomposition and stability are basically the same as a or b. The detailed analysis of 9 combination are listed in Table 6.

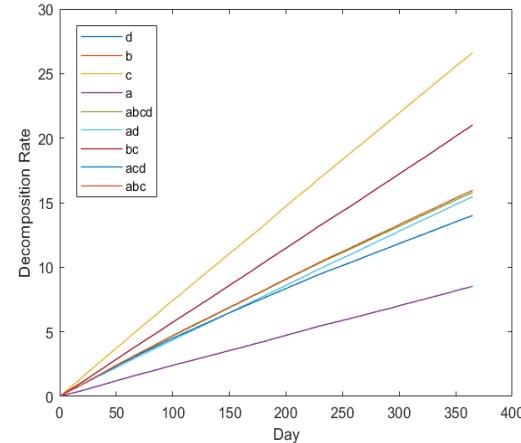
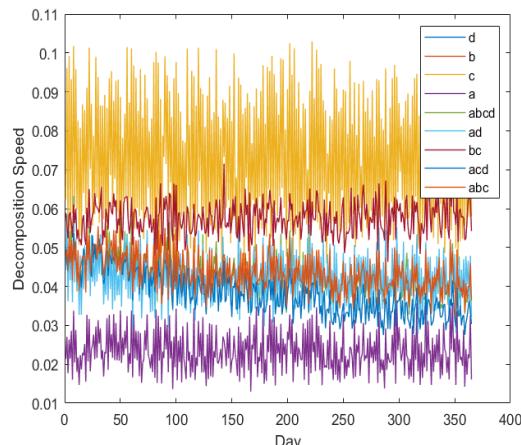
Accordingly, we consider 9 combinations of fungi and 2 environments.

**Table 6. Combinations / Advantages and Disadvantages**

Combinations	Advantages and Disadvantages
a or b	Fast decomposition rate, weak adaptability
c or d	Slow decomposition rate, strong adaptability
b and c	The initial decomposition rate and adaptability are good, and eventually tend to c
a and d	The initial decomposition rate and adaptability are good, and eventually tend to a
a, b and c	Always maintain a good decomposition rate and stability, final decomposition rate is slightly better than the stability
a, c and d	Always maintain a good decomposition rate and stability, the final stability is slightly better than the decomposition rate
a, b, c and d	decomposition rate and stability have always been kept in good balance



(a) Decomposition Speed in Arboreal (b) Decomposition Rate in Arboreal



(c) Decomposition Speed in Arid (d) Decomposition Rate in Arid

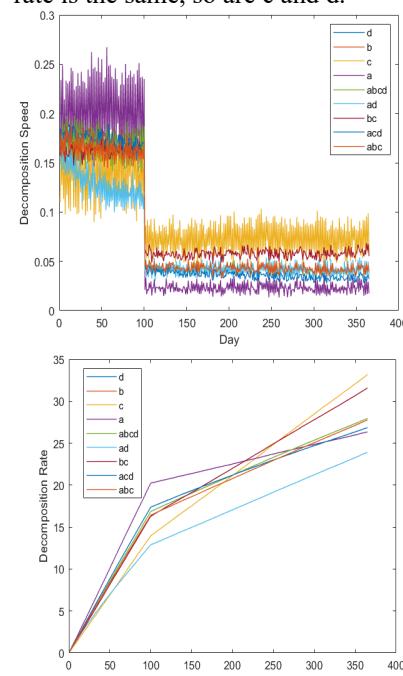
Figure 6. Decomposition Speed and Decomposition Rate

Figure 6 shows our simulation results. We can see that in Arboreal and Arid, the decomposition rate of fungi is quite different. In Arboreal area, due to suitable environment, competition become the only impact factor of decompose, so the decomposition rate can reach 70. While in Arid area, the precipitation is too low, so the decomposition rate only reaches 30.

#### 4.3 THE IMPORTANCE OF BIODIVERSITY

At the beginning of the experiment, the fungi was allowed to grow under suitable conditions, and the environment became very unsuitable for fungi growth on the one hundredth day (from arboreal to arid environment).

As shown in Figure 7(a), after 100 days, when environment changed dramatically, all fungus's decomposition speed drops, but the decline is different. In Figure 7(b) we can see that the decomposition rate (sum of decomposition speed) of nine fungi combinations. Note that when cultivated separately, a and b's decomposition rate is the same, so are c and d.



(a) Decomposition Speed Changes Over Days (b)

### Decomposition Rate Changes Over Days

#### Figure 7. Results of Experiment

Analyze the decomposition rate of abcd, it can be seen that during 365 days, its decomposition rate is in middle of all nine groups, which indicated the pros and cons of biodiversity. Group a and b decompose very fast in the first 100 days, but their decomposition rate drops a lot when environment changes. We can predict that if the environment does not recover, they will have the lowest decompose rate.

#### 5. CONCLUSION

In this article, we propose a model that can predict the decomposition rate of different combinations of fungi in different environments. We first use certain rules to classify fungi, and use k-means method to verify the correctness of our classification. Classification provides the basis for us to use the Markov chain model. Then we use the Markov absorption chain model to simulate the decomposition of fungi based on the relevant knowledge of the ecological niche, and predict the population density of fungi in the long and short term.

When describing the influence of fungi on the environment, we put forward the concept of decomposition coefficient. We take temperature and moisture as the main evaluation indicators, and normalize the temperature and humidity according to the decomposition of the fungus, and then use a normal distribution model and an experience-based equation to describe the effect of changing environment on the decomposition rate of fungi.

In short, this model is suitable for large-scale prediction of fungal decomposition rate, and can help select a combination of fungi with advantages in a certain environment, or select an environment suitable for fungal growth, and provide suggestions for improving fungal decomposition rate. At the same time, this model can verify the importance of biodiversity and can be used for environment recovery.

#### ACKNOWLEDGMENT

This study was funded by "First-class Course" of Anhui University of Finance and Economics (acylk202008), and the Teaching and Research Fund Project of the Anhui University of Finance and Economics (acjyyb2020011 and acjyyb2020014).

#### REFERENCES

- [1] G. M. Gadd. Geomycology: biogeochemical transformations of rocks, minerals, metals and radionuclides by fungi, bioweathering and bioremediation. *Mycol. Res.*, 2007, 111(1): 3-49.
- [2] N. Lustenhouwer, D. S. Maynard, M. A. Bradford, et al. A trait-based understanding of wood decomposition by fungi. *Proc. Natl. Acad. Sci. U. S. A.*, 2020, 117(21): 11551-11558.
- [3] A. Rapaport and M. Veruete. A new proof of the competitive exclusion principle in the chemostat. *Discrete Continuous Dyn. Syst. Ser. B*, 2019, 24(8): 3755-3764.
- [4] J. Cortez. Field decomposition of leaf litters: relationships between decomposition rates and soil moisture, soil temperature and earthworm activity. *Soil Biol. Biochem.*, 1998, 30(6): 783-793.
- [5] M. J. Fuhr, C. Stührk, M. Schubert et al. Modelling the effect of environmental factors on the hyphal growth of the basidiomycete *Physisporinus vitreus*. *J. Basic Microbiol.*, 2012, 52(5): 523-530.
- [6] Y. H. Fang, H. S. Chen, Y. Lin et al. Classification of Northeast China Cold Vortex Activity Paths in Early Summer Based on K-means Clustering and Their Climate Impact. *Adv. Atmos. Sci.*, 2021, 38(3): 400-412.
- [7] D. Arthur and S. Vassilvitskii. k-means++: the advantages of careful seeding. In Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, USA, January 2007: 1027-1035.
- [8] Bo. Lindqvist. How Fast Does a Markov Chain Forget the Initial State? A Decision Theoretical Approach. *Scand. J. Stat.*, 1977, 4(4): 145-152.
- [9] P. K. Pollett. The determination of quasistationary distributions directly from the transition rates of an absorbing Markov chain. *Math Comput. Model.*, 1995, 22(10-12): 279-287.
- [10] Y. L. Xi and Atsushi Hagiwara. Competition between the Rotifer *Brachionus calyciflorus* and the Cladoceran *Moina macrocopa* in Relation to Algal Food Concentration and Initial Rotifer Population Density. *J. Freshwater Ecol.*, 2007, 22(3): 421-427.
- [11] J. F. Liao, L. N. Tang, G. F. Zhang et al. A neighbor decay cellular automata approach for simulating urban expansion based on particle swarm intelligence. *Int. J. Geogr. Inf. Sci.*, 2014, 28(4): 720-738.
- [12] L. K. Mühlbauer, M. Schulze, W. S. Harpole et al. gauseR: Simple methods for fitting Lotka-Volterra models describing Gause's 'Struggle for Existence'. *Ecol. Evol.*, 2020, 10(23): 13275-13283.
- [13] C. Krishna. Solid-state fermentation systems—an overview. *CritRev. Biotechnol.*, 2005, 25(1-2): 1-30.
- [14] E. Kerry. Effects of temperature on growth rates of fungi from subantarctic Macquarie Island and Casey, Antarctica. *Polar Biol.*, 1990, 10(4): 293-299.
- [15] K. Chen, W.G. Yao, Z. Zhang et al. Establishment risk analysis and divisions of TCK in China. *Acta. Phys. Sinch. Ed.*, 2002, 32(4): 312-318.
- [16] L.P. Hou, P. He, X.S. Pan et al. A review on the methods of determining ecological thresholds. *J. Appl. Ecol.*, 2021, 32(2): 711-718.
- [17] A. N. Beştaş et al. Double Gaussian distribution of barrier height for FeCrNiC alloy Schottky contacts on p-Si substrates. *Appl. Surf. Sci.*, 2014, 318(1): 280-284.
- [18] F. T. de Vries, R. I. Griffiths, M. Bailey et al. Soil bacterial networks are less stable under drought than fungal networks. *Nat. Commun.*, 2018, 9(1): 3033.
- [19] G. Barcenas-moreno, M. Gomez-brandon, J. Rousk et al. Adaptation of soil microbial communities to temperature: comparison of fungi and bacteria in a laboratory experiment. *Glob. Chang. Biol.*, 2009, 15(12): 2950-2957.
- [20] J. Sarkar, J. R. Chicholikar and L. S. Rathore. Predicting Future Changes in Temperature and Precipitation in Arid Climate of Kutch, Gujarat: Analyses Based on LARS-WG Model. *Curr. Sci.*, 2015, 109(11): 2084-2093.

- [21] Karine G.D. Lopes and Júlio César Bicca-Marques, (Callicebus bernhardi). *J. Therm. Biol.*, 2017, 69: 104-109.  
“Ambient temperature and humidity modulate the behavioural thermoregulation of a small arboreal mammal

# Evaluation Model of Minors' mental Health Based on Improved Factor Analysis

Xintao Cui<sup>1\*</sup>, Jianglong Mao<sup>2</sup>, Junfang Zhang<sup>1</sup>

<sup>1</sup>School of Metallurgy and Energy, North China University of Science and Technology, Tangshan, 063210, China;

<sup>2</sup>School of Materials Science and Engineering, North China University of Science and Technology, Tangshan, 063210, China

\*Corresponding Author.

**Abstract:** With the rapid development of China's economy and society, China has paid more and more attention to the mental health of minors, and a simple and effective model is urgently needed to evaluate the mental health level of minors. Based on this, this paper analyzes the physiological and psychological development, the increase of competitive pressure, the expansion of social experience and the change of thinking mode, and establishes a set of mental health assessment model for minors, which can be used to reasonably, accurately and quickly assess the mental health status of all minors. First, preprocess the data to remove the influence of irrelevant data and missing data, and standardize the preprocessed data. Then, through factor analysis, we turn the 10 evaluation indicators given in the annex into three comprehensive indicators: physiological and psychological development, increasing competitive pressure, expanding social experience and changing thinking mode. Then, we calculate the TOPSIS score and fit the regression equation with the TOPSIS score as the dependent variable and the comprehensive index as the independent variable. Finally, an evaluation equation applicable to all minors' mental health is established, and the mental health of minors can be evaluated by combining the evaluation criteria.

**Keywords:** Factor analysis; Multiple linear regression; TOPSIS score; Mental health evaluation.

## 1. INTRODUCTION

With the rapid development of China's society and economy, the competitive pressure is increasing, and the minors are in an important period of physical and mental development. Their own social experience has expanded and their way of thinking has changed, resulting in various psychological problems in their study, life and interpersonal communication<sup>[1-4]</sup>. Therefore, it is of great significance to construct the evaluation system of minors' mental health.

## 2. EXTRACTION OF PRINCIPAL COMPONENT INDEX BY FACTOR ANALYSIS

In order to avoid data redundancy caused by information duplication among indicators, this paper uses factor analysis to extract comprehensive indicators for later analysis and processing.

### 2.1-KMO, BARTLETT SPHERICAL INSPECTION

In this paper, KMO and Bartlett are used to test the original data to verify whether the original data can use factor analysis mode<sup>[5-6]</sup>. The test results show that the original

data can be used for factor analysis within 95% confidence interval.

Table 1.KMO and bartlett inspection table.

KMO and bartlett inspection.	
KMO sampling suitability quantity.	0.721
Bartlett sphericity test	Approximate square freedom significance
	chi- 1655.163 45 0.000

### 2.2 ESTABLISHMENT OF R-TYPE FACTORMODEL

The relationship between 10 indexes and common factors is constructed, and the following determinants are obtained:

$$\begin{cases} x_1 = u_1 + a_{11}f_1 + a_{12}f_2 + \dots + a_{1m}f_m + \varepsilon_1 \\ x_2 = u_2 + a_{21}f_1 + a_{22}f_2 + \dots + a_{2m}f_m + \varepsilon_2 \\ \vdots \\ x_{10} = u_{10} + a_{101}f_1 + a_{102}f_2 + \dots + a_{10m}f_m + \varepsilon_{10} \end{cases}$$

$a_{ij}$  represents an original variable  $x_i$  Common factor  $f_j$  The covariance between, namely:  $a_{ij} = \text{cov}(x_i, f_j)$ . Construct a matrix from the determinant. The whole determinant can be abbreviated as:  $x = u + Af + \varepsilon$ , among  $f = (f_1, f_2, \dots, f_m)$  ( $m \leq 10$ ) Is a common factor variable,  $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_8)$  is a special factor vector.  $A_{8 \times m} = (a_{ij})$  It is called factor load matrix, and the rank of matrix A is assumed to be m.

By calculating the sum of squares of row elements of A matrix respectively.  $h_i^2$  And that sum of square of column elements  $g_j^2$ , solve the original variables.  $x_i$  The dependence on the common factor and the contribution of the common factor  $f_j$  to the overall index data X, and then select the most important comprehensive factor. Where the sum of squares of row elements and column elements is respectively:

$$h_i^2 = \sum_{j=1}^m a_{ij}^2 g_j^2 = \sum_{i=1}^p a_{ij}^2$$

### 2.3 DETERMINE THE NUMBER OF FACTORS

In this paper, three common factors are extracted from the turning point of the gravel map and the cumulative contribution rate of variance interpretation table. The variance contribution rates of the extracted three common factors are 76.735%, 12.021% and 3.502%, respectively. The cumulative variance contribution rate of the first three common factors has reached 92.258%, that is, the first

three common factors can contain 92.258% of the information of the original index, so the first three common factors are selected to reflect the whole.

#### 2.4 FACTOR SCORE

Factor analysis is to express variables as linear combinations of common factors and special factors. In this paper, we can express common factors as linear combinations of original variables in reverse, that is, we can get factor scores, and the generated component score coefficient matrix is shown in table.

Table 2 Matrix table.

influencing factor	component part		
	1	2	3
interpersonal relationship	0.634	-0.256	0.077
stubbornly biased	0.422	-0.333	0.596
Obsessive symptoms	0.245	-0.333	0.452
Obsessive symptoms	0.457	-0.295	0.103
depressed	0.652	-0.259	-0.428
anxious	0.658	-0.179	-0.451
Learning pressure	0.487	0.404	-0.151
Psychological imbalance	0.265	0.458	0.388
Emotional imbalance	0.558	0.409	0.036
maladjustment	0.482	0.538	0.141

From the above table, we can get the expressions of three principal components:

$$f_1 = 0.634x_1 + 0.422x_2 + \dots + 0.482x_{10}$$

$$f_2 = -0.256x_1 - 0.333x_2 + \dots + 0.538x_{10}$$

$$f_3 = 0.077x_1 + 0.596x_2 + \dots + 0.141x_{10}$$

The positive load coefficient of depression and anxiety in first principal component is large, so first principal component is called the increase of competitive pressure. The positive load coefficient of psychological imbalance and maladjustment in the second principal component f2 is larger, so the second principal component is called physiological and psychological development and development. The positive load coefficient of paranoia and obsessive-compulsive symptoms in the third principal component f3 is large, so the third principal component is called the expansion of social experience and the change of thinking mode. The three principal components reflect the influencing factors of mental health of minors more intuitively, and the data of the three principal components are obtained according to the above expressions.

#### 3. CALCULATION OF TOPISIS SCORE

Through the above analysis, three principal component indicators are obtained, among which the policy support and national attention are the maximum indicators, and the pressure of college students is the minimum indicators. In this paper, a scoring model based on TOPISIS is established<sup>[7-8]</sup>.

#### 3.1 CALCULATE THE SCORE AND NORMALIZE IT.

Finally, the score is calculated and normalized, and the maximum value is defined as:

$$z^+ = (z_1^+, z_2^+, \dots, z_5^+) = (\max\{z_{11}, z_{21}, \dots, z_{n1}\}, \dots, \max\{z_{15}, z_{25}, \dots, z_{n5}\})$$

The minimum value is defined as:

$$z^- = (z_1^-, z_2^-, \dots, z_5^-) = (\min\{z_{11}, z_{21}, \dots, z_{n1}\}, \dots, \min\{z_{15}, z_{25}, \dots, z_{n5}\})$$

Define the first  $i (i=1, 2, \dots, n)$  the distance between each evaluation object and the maximum value and the minimum value is:

$$D_i^+ = \sqrt{\sum_{j=i}^5 w_j (z_j^+ - z_{ij})^2} \quad D_i^- = \sqrt{\sum_{j=i}^5 w_j (z_j^- - z_{ij})^2}$$

The TOPSIS score corresponding to the mental health status of minors is:

$$S_i = \frac{D_i^-}{D_i^+ + D_i^-}$$

#### 4. EVALUATION MODEL BASED ON MULTIPLE LINEAR REGRESSION

The above three principal component indicators are used as independent variables, and the TOPSIS score of mental health status of minors is used as dependent variable. Through analysis, it can be seen that there is a linear correlation between independent variables and dependent variables. The OLS general least square method is used to solve the correlation coefficient of each index, so as to obtain the multiple regression equation with three principal component indexes<sup>[9-10]</sup>.

#### 4.1 TEST OF SIGNIFICANCE

To test whether there is a real linear relationship between the dependent variable and the independent variable, the significance test is carried out first, and the results are shown in the following table:

Table 3 Regression equation fitting test table.

model	R	R square	Adjusted R square.	Error standard estimation.	P
1	0.963a	0.926	0.926	0.221	0.0008

a. Predicted variables: (constant), component 1, component 2 and component 3.  
b. dependent variable: TOPSIS value.

It can be seen from the above table that the standard P=0.0008<0.05 corresponding to f test means that the regression equation is considered meaningful within 95% confidence interval. Among them, R2 and R2adjusted are both close to 1, which indicates that the regression model has high fitting goodness.

#### 4.2 USING MULTIPLE LINEAR REGRESSION TO FIND THE EVALUATION EQUATION.

Table 4 Coefficient table of multivariate regression equation.

model	Unstandardized coefficient		
	B	Standard error	significance
1	(constant) 12.860	0.044	0.000
	Composition 1 -0.669	0.005	0.000
2	Composition 2 -0.001	0.010	0.000
3	Composition 3 -0.265	0.011	0.000

In this paper, multiple linear regression is used to predict each index coefficient. f1, f2, f3 are independent variables, Y is dependent variable, and satisfy the following linear relationship:

$$y = 12.8605 - 0.6690 * f_1 - 0.0011 * f_2 - 0.2646 * f_3$$

Sensitivity analysis shows that the scores obtained by the evaluation system are basically consistent with TOPSIS scores, and the change trend of scores is basically

consistent. Therefore, the evaluation model established in this paper is verified to be accurate.

#### 5.MODEL EXTENSION

This model is applicable to a wide range of fields, not only for the establishment of the mental health system of minors under the influence of multi-dimensional factors, but also for a variety of evaluation systems, such as the adolescent health evaluation system and the elderly health evaluation system. Through factor analysis, various influencing factors are extracted into several concise factors, and the data obtained by TOPSIS is more convincing. It can quickly, accurately and simply evaluate different objects reasonably.

#### 6.MODEL EVALUATION

##### 6.1EVALUATION OF THE MODEL

(1) We have considered 10 indicators that affect teenagers' mental health, obtained the correlation of the 10 indicators given in the title by principal component analysis, and divided the factors that affect teenagers' mental health into three main components: first, the increase of competitive pressure; second, the development of physiology and psychology; third, the expansion of social experience and the change of thinking mode.

(2) In order to better evaluate the mental health status of teenagers, we use topsis solution distance method to evaluate the mental health of teenagers in this area.

(3) In order to get the linear relationship between adolescents' mental health scores and three principal components, we use multiple linear regression analysis model to evaluate adolescents' mental health.

(4) By establishing the model, we tested the significance of the evaluation model of adolescent mental health established by ourselves, and compared the model score with the actual score, and found that the fitting effect between the model score and the actual score was very good.

##### 6.2EVALUATION OF ALGORITHM BASED ON

(1) The evaluation of TOPSIS solution distance method: TOPSIS can make full use of the information of original data, and the results can accurately reflect the gap between evaluation schemes. The basic process is to unify the index types of the original data matrix (general normalization processing) to get the normalized matrix, then standardize the normalized matrix to eliminate the influence of each index dimension, and find the best scheme and the worst scheme in the limited scheme, then calculate the distances between each evaluation object and the best scheme and the worst scheme respectively, and obtain the relative closeness between each evaluation object and the best scheme, which is used as the basis for evaluating the pros and cons. This method has no strict restrictions on data distribution and sample content, and data calculation is simple and easy.

(2) Evaluation of principal component analysis: Principal component analysis is a dimensionality reduction algorithm, which can convert multiple indicators into a few principal components, which are linear combinations of original variables and have no correlation with each other, and can reflect most information of original data. The defect lies in that when the sign of factor load of

principal component is positive or negative, the meaning of comprehensive evaluation function is not clear and the naming clarity is low.

(3) Evaluation of multiple linear regression analysis: Multiple linear regression is used to explore the relationship between multiple independent variables and dependent variables in a phenomenon, and the optimal combination of multiple independent variables can jointly predict or estimate the dependent variables, which is more effective and more realistic than using only one independent variable to predict or estimate. Sometimes, in regression analysis, it is only a guess which factor to choose and which expression to use, which affects the diversity of factors and the unpredictability of some factors, and makes regression analysis limited in some cases.

##### 6.3ADVANTAGES OF THIS MODEL IN COMPREHENSIVE EVALUATION

(1) Using the principal component analysis method, 10 evaluation indexes are divided into three categories to reduce the evaluation dimension, and then taking these three principal components as independent variables and TOPSIS scores as dependent variables, a multiple linear regression equation is established, which can better reflect the mental health of teenagers.

(2) When we set up the model, we set the parameters to be easy to adjust to meet the needs of further research, and further explained the sensitivity of parameters and the robustness of the model through sensitivity analysis.

#### 7.SUGGESTION

We set up a mathematical model to process and analyze the data of the factors affecting the mental health of minors, and get the following results:

Compulsive symptoms, study pressure and emotional imbalance are the main factors that affect the mental health of minors. No matter from the perspective of boys or girls, obsessive-compulsive symptoms, study pressure and emotional imbalance are the three main factors that affect the mental health of boys and girls, but these three factors affect the mental health of boys to a greater extent than that of girls. The reason is that boys are naturally playful, obsessed with games, greatly influenced by the network environment, and boys in adolescence are rebellious, paranoid, easily excited and have intense emotional reactions. Compared with girls, boys have to bear greater social pressure, which leads to constant pressure from the outside world on boys' study. The high pressure of study has a serious impact on their mental health.

It can be seen that the current mental health of minors is not optimistic, and the unhealthy psychological harm is very great, which will endanger students' personal growth and development, family happiness and social stability. Therefore, we give the following suggestions through investigation and data analysis:

**Family:** Let go of love, approach children, be good at enlightening and listening

Parents should not give their children too many rigid rules and regulations, nor impose too high expectations on them. Children should be encouraged to participate in collective

activities and cultural and sports activities, cultivate hobbies, and let children build new excitement; Be good at listening to children's thoughts, truly understand their hearts and give them the right to choose freely; At the same time, parents should set an example, take a natural attitude towards life, learn positive ways to deal with various pressures, and don't shy away from difficulties. Children's exposure to positive and optimistic family environment will be affected, which will help cultivate children's quality of bearing pressure and frustration.

**School:** Respect individual differences and implement quality education

Schools should regularly arrange students' psychological lectures, carry out mental health education and provide relevant psychological counseling to students; Schools should train teachers in psychological knowledge, "A good word warms the winter in three months, and a bad word hurts the cold in June." Teachers are groups that are in direct contact with students. Teachers' words and deeds will directly affect students' inner thoughts. Teachers' quality should be strengthened. Schools should build a warm, harmonious and beautiful campus, respect individual differences, and never form an atmosphere of "only points theory". Achievements can represent many things, but not everything. We should promote the all-round development of students' morality, intelligence, physique, beauty and labor, and cultivate social talents with excellent comprehensive quality.

## REFERENCES

- [1] Song Tieli. Strategies for cultivating children's healthy psychology. Tonghua, Jilin, Department of Education, tonghuanormal University 134002. Journal of tonghua normal University. 2005.(03).26-27.
- [2] Bo Huang. Optimization strategy of mental health education in primary schools. School of Educational Sciences, Xianyang Normal University. Asia-PacificEducation. 2019.(12).151.
- [3] Chen Dan, Quan Zhixing, Ai Mengyao, Zong Chunshan, Xu Jiannong. Influencing factors of adolescents' mental health. Beijing Youth Legal and Psychological Counseling Service Center. Experimental Middle School of Chengdu University of Electronic Science and Technology, Sichuan Province. Chinese Journal of Health Psychology.2020.28 (09).1402-1409.
- [4] Li Juan, Yin Shishu, Gao Yanwen, LiDongmei. Study on influencing factors of adolescent mental health based on canonical correlation. School of Finance, Anhui University of Finance and Economics. anhui university of finance Schoolof Management Science and Engineering. Journal of Fuyang Teachers College (Natural Science Edition).2016.33(04).93-97.
- [5] Xiao Yang, Spencer Pan, Hu Xudong, Wu Jiang, Yvonne Wang. Evaluation of power emergency drill based on factor analysis. Power safety technology. Suzhou Power Supply Branch of State Grid Jiangsu Electric Power Co., Ltd.2020.22(06), 40-44.
- [6] Jiang Wei, Xie Xueping, Liao Jinchun, Zhang Yufei. Study on the economic competitiveness of major cities in Guangxi based on factor analysis. Personnel Department of Hezhou University, School of Mathematics and Computer Science of Hezhou University. modern business trade industry.2019.40 (23): 11-1.
- [7] Liu Fei, Gong Ting. Comprehensive evaluation of high-quality development in Hubei Province based on entropy Topsis model. Law School of Huazhong University of Science and Technology. Hubei Elvtou Real Estate Group Co., Ltd. statistics and decision.2021. (11).85-88.
- [8] Chen Zhixia, Jeff. Evaluation of urban happiness index based on TOPSIS and grey relational analysis. School of Public Administration, Huazhong University of Science and Technology. statisticsand decision.2021.37 (09).59-62.
- [9] Guan Yongqian. Influencing factors of GDP in Henan province based on multiple linear regression. Guizhou University of Finance and Economics. Rural Economy and Technology. 2021.32.(05).221-224.
- [10] Wang Yiran, Wei Jia, Fan Siren. Analysis of factors influencing the development of high-tech enterprises based on multiple linear equations. Torch High-tech Industry Development Center, Ministryof Science and Technology. bulletin of science and technology.2020.36 (12).112-115.

# Credit Rating Classification Model Based on AHP-EWM

Yaxun Dai<sup>1, 2\*</sup>, Bingbing Shi<sup>2</sup>, Fanjie Jin<sup>3</sup>

<sup>1</sup>Engineering Computing and Simulation Innovation Lab, North China University of Science and Technology, Tangshan 063210, China;

<sup>2</sup>School of Artificial Intelligence, North China University of Science and Technology, Tangshan 063210, China;

<sup>3</sup>School of Mechanical Engineering, North China University of Science and Technology Tangshan 063210, China;

\*Corresponding Author.

**Abstract:** The potential personal credit risk is an important hidden danger that leads to the bank crisis. This paper mainly solves this problem and studies the personal credit risk assessment of bank customers, selects risk assessment indicators and establishes a high precision assessment model, so as to maximize the bank credit income. After preprocessing of the indicators data through the system clustering original 25 indicators fall into four general index, respectively, for the assets, liabilities, credit history and personal information, thus established the new personal credit risk evaluation index system, and then through the analytic hierarchy process (ahp) and entropy weight method, linear combination method to calculate the weights of each indicator, Then, the weighted Topsis model is used to score and estimate the customer's credit, and k-means algorithm is used to cluster the credit score. According to the clustering results, the corresponding ABCD four grades are divided, and the credit good degree decreases successively. Among them, the customer's credit score is divided into grade A between [71.5, 100]. It is classified as grade B between [57.6, 71.5], grade C between [50.2, 57.6] and grade D between [10, 50.2].

**Key words:** K-means clustering; Weighted Topsis; Hierarchical analysis; Entropy weight method

## 1. PROBLEM ANALYSIS

In the absence of customer credit rating, the establishment of a model to classify customer credit rating is actually to quantify the various indicators of customers. We build an individual credit risk evaluation index system, reduce the dimension of indicators through system clustering, and establish a new evaluation index system. Using the method of analytic hierarchy process (ahp) and entropy weight method to determine the weight of each index, and then USES the weighted Topsis model to evaluate the model, get each customer's credit score, in order to classify the customer's credit rating, we use the K - means algorithm to cluster of customer credit score, the result is divided into four classes, Therefore, the customer's credit score corresponds to four credit grades ABCD respectively.

## 2. PERSONAL CREDIT RISK ASSESSMENT MODEL BASED ON K-MEANS CLUSTERING

To establish model in the case of category unknown credit rating classification with the customers, will issue one to extract the 25 indexes is one of the important reference in real life personal credit evaluation data, clustering and

combining the actual situation, so the system will be 25 indicators divided into four kinds of general indicators, respectively, for the assets, liabilities, credit history, and personal information. Then, we choose the linear combination of analytic hierarchy Process and entropy weight method to calculate the weight of these indicators, and then use Topsis method to calculate the credit score and valuation of each customer. Finally, cluster according to the score and valuation, and divide the credit rating of users into four categories: ABCD.

### 2.1 SYSTEM CLUSTERING

System clustering method which is at the beginning of each sample as a kind, then according to the distance criterion, the relative near samples were first divided into small classes, and then have the distance between the aggregation of small class according to their class to merge, constantly continue, until obtain appropriate classification requirements, which can easily find out the main factors influencing the system.

Firstly, we obtained that personal credit evaluation is mainly affected by assets, liabilities, historical credit and personal information by referring to relevant materials and combining with the actual situation. Then, we used factors to measure the correlation coefficient of these four aspects. In the cluster analysis of factors (specific factors), the first step is to determine the correlation coefficient of factors in different classes, the specific method is as follows: We define two factors,  $x_i$  and  $y_i$ , that correspond to our specific factor and type. And calculate its correlation coefficient, the formula is as follows:

$$r = \frac{\sum_{j=1}^n (x_{ji} - \bar{x}_i)(y_{ji} - \bar{y}_i)}{\sqrt{\sum_{j=1}^n (x_{ji} - \bar{x}_i)^2} \sqrt{\sum_{j=1}^n (y_{ji} - \bar{y}_i)^2}}$$

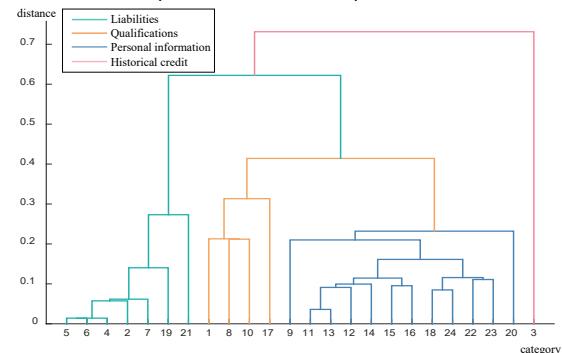


FIG. 1 System clustering diagram

Secondly, Euclidean distance is used as the distance

calculation criterion. According to the calculated distance between each class, the two nearest classes are merged into a new class, and then the distance between the new class and other classes is calculated. Similarly, the two nearest classes are merged into a new class according to the calculated distance. The above process is repeated, and finally the 25 factors are grouped together. Finally, each factor was labeled according to the order of the table, and the correlation coefficient matrix was taken as the input parameter. After clustering analysis by MATLAB, the cluster diagram of the system was obtained, as shown in Figure 1.

According to the clustering results, the 25 indicators can be divided into 4 categories of general indicators, among which:

- Assets include current deposit status A1, savings account/bonds A6, percentage of bonds in disposable income A8, property status A12;
- Liabilities include loan term A2, loan purpose A41/A42/A43, loan amount A5, other installment plan A14, current number of loans with the bank A16;
- Personal information includes working years A7, gender and marital status A91/A92/A93/A94, other debtors / guarantors A10, current residence years A11, age A13, housing status A15, occupation type A17, telephone number A19, foreign worker A20;
- Historical credit includes credit history A3.

## 2.2 WEIGHT DETERMINATION BASED ON AHP-EWM

After four indicators are established, ahp and entropy weight method are used to determine the weight of each indicator. The specific algorithm steps are as follows:

### (1) Ahp determines the weight

The judgment matrix is obtained according to the importance of assets  $C_1$ , liabilities  $C_2$ , historical credit  $C_3$  and personal information  $C_4$  to the personal credit P of the target layer and the related importance of the judgment quasi-side layer.

Table 1 Judgment matrix

P	$C_1$	$C_2$	$C_3$	$C_4$
	1.0	3	2	5
	0.3	1	0.25	2
	0.5	4	1	5
	0.2	0.5	0.2	1

According to the above judgment matrix, consistency test was carried out and  $CR=0.0568$  was calculated to be less than 0.1. It can be considered that the consistency of the judgment matrix is within our acceptable range.

In order to ensure the robustness of the results, arithmetic average method, geometric average method and eigenvalue method are used to calculate the weights and average values respectively, so as to avoid the deviation caused by using a single method, and the results are as follows:

Table 2 Weight calculation for the three results

$C_i$	Arithmetic mean	Geometric mean	Eigenvalue	$w_{1i}$
$C_1$	0.314	0.318	0.315	0.316
$C_2$	0.325	0.321	0.326	0.324
$C_3$	0.119	0.115	0.111	0.115
$C_4$	0.232	0.236	0.240	0.233

### (2) Combined with entropy weight method to determine the comprehensive weight

The weight determined by AHP method is subjective. In order to make the weight of the evaluation index more reasonable and reliable, we determine the comprehensive weight by combining the entropy weight method, which is based on the index data and calculates the entropy by deducing and calculating the entropy.

Attachment to all the customer's credit rating as the evaluation sample, including assets, liabilities, quality, and individual conditions, four evaluation indexes, is composed of matrix, and the standardization of each element to judge whether there is a person in the negative Z matrix, the existence of a negative number,  $Z'$  matrix matrix standardized processing, formula for standardization; Calculate the weight of the  $i$ th index of the JTH index and regard it as the probability used in relative entropy calculation. Calculate the information entropy of the JTH index according to the following formula:

$$e_j = -\frac{1}{\ln n} \sum_{i=1}^n P_{ij} \ln(P_{ij}), (j = 1, 2, \dots, m)$$

In addition, the information entropy redundancy is calculated. Finally, the information utility value is normalized to obtain the entropy weight of each evaluation index, and the formula is as follows:

$$b_j = \frac{d_j}{\sum_{j=1}^m d_j}, j = 1, 2, \dots, m$$

The weight of each indicator is calculated as follows:

Table 3 Weight results of each index calculated by entropy weight method

	$C_1$	$C_2$	$C_3$	$C_4$
$w_{2i}$	0.31579	0.327861	0.116345	0.22576

### (3) Combine the two to determine the weight

In order to combine ahp and entropy weight method, the average offset distance D is defined. By collecting a large number of relevant psychology literature, found in the personal credit assessment is divided into four levels of screening, we find out the four general indexes as secondary screening index, the situation of these assets directly reflects the loan to the customer if there is a stable economic conditions and enough ability to repay, is an important premise and guarantee for the credit can be recycled; Liabilities can also reflect the current asset status and solvency of individuals to a certain extent. Credit record is a true reflection of a person's character, and it is an objective basis to judge whether a customer will perform the contract when lending with other institutions or individuals. Therefore, historical credit also plays an important role in personal credit evaluation. Personal information is an auxiliary assessment of a customer's economic strength. Therefore, we obtained the authoritative ranking of the importance degree of the 4 indicators, and obtained their relevant weights according to the ranking, as shown in the following table:

Table 4 Authority ranking of general data

$C_i$	$J_i$	formula	$W_j$
$C_1$	1		0.295
$C_2$	2		0.263

$C_3$	3	$W_j = (1 - \frac{J_i}{\sum_{k=1}^3 J_k})/2$	0.240
$C_4$	4		0.212

The average offset distance D is defined by the one-dimensional distance formula between two points, which is as follows:

$$D_x = \sum_i^3 |W_j - w_{xj}|, x = 1, 2$$

The value of x for 1, 2, on behalf of the meaning of analytic hierarchy process  $w_{1j}, w_{2j}$  and entropy weight method represents the ith a weight, so  $|W_j - w_{xj}|$  represents authority results and calculation results, the difference between  $D_x$  for x a evaluation model of the weight distribution, The main results  $D_1 = 0.093965, D_2 = 0.082$ .

According to the above results, it can be concluded that the error of entropy weight method is smaller than that of analytic hierarchy process. In order to comprehensively consider the influence of subjective and objective factors, linear weighting method is adopted:

$$w_j = \alpha w_{1j} + (1 - \alpha)w_{2j}, 0 < \alpha < 1$$

Where  $w_{1j}$  represents the weight of the four indexes in the ANALYTIC hierarchy process,  $w_{2j}$  represents the weight matrix in the entropy weight method,  $w_{2j}$  represents the index weight of individual credit evaluation after the combination of the two methods, and  $\alpha$  is the weight factor. In order to determine the weight factor, we consider the average offset distance. Because a smaller average offset distance means more accurate results, we assign a higher weight to the algorithm with a smaller deviation, whose formula  $\alpha$  is as follows:

$$\alpha = \frac{D_2}{D_1 + D_2}$$

The weight assigned by the analytic hierarchy process is small, while the weight assigned by the entropy weight method is large. The main result data are shown in the following table:

Table 10 Algorithm combined with data table

	AHP	EWM	Combing
$C_1$	0.316	0.3157934	0.315
$C_2$	0.324	0.32786136	0.326
$C_3$	0.115	0.11634524	0.115
$C_4$	0.233	0.2257647	0.227

In this way, through the linear combination of the two evaluation algorithms, we get the final weight of assets, liabilities, historical credit and personal information, thus obtaining the personal credit evaluation system.

### 2.3 CALCULATE CREDIT SCORES ACCORDING TO TOPSIS

After the weight of the four indicators was calculated by AHP-EWM, the weighted Topsis model was used to evaluate the indicators and the individual credit score of each customer was obtained. The specific algorithm steps are as follows:

Step1: Firstly, all the data involved in the selected indicator are processed forward to unify the indicator type, and the forward matrix is obtained. Among them, assets, historical credit and personal information are extremely large, while liabilities are extremely small. The reciprocal method ( $\frac{1}{x}$ ) is used for extremely large processing of this

index, and the polar consistency matrix  $R = (r_{i1}, r_{i2}, r_{i3}, r_{i4})$  of the evaluation index is obtained.

Step2: Use AHP-EWM to determine the index weight vector

Step3: standardize the forward transformation matrix obtained by step1 to eliminate the influence of different index dimensions, namely

$$z_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^n x_{ij}^2}}$$

Step4: according to the step to get on the standardization of the matrix maximum defined  $Z^+ = \max|z_{ij}|$  and minimum  $Z^- = \min|z_{ij}|$ , which is suitable for the optimal vector  $Z^+, Z^-$  as the vector of inferior quality.

Step5: define the distance between the ith evaluation object and the maximum value  $D_i^+ = \sqrt{\sum_{j=1}^m w_j (Z_j^+ - z_{ij})^2}$  and minimum distance  $D_i^- = \sqrt{\sum_{j=1}^m w_j (Z_j^- - z_{ij})^2}$ .

Step6: Finally calculate the score before normalization of the ith evaluation index

$$S_i = \frac{D_i^-}{D_i^- + D_i^+}$$

At this point, the larger the various evaluation indexes are, the closer the state of the evaluated object is to  $D_i^-$ , and the greater the comprehensive evaluation value is, the higher the individual credit score will be. On the contrary, it is close to  $D_i^+$ , and the smaller the comprehensive evaluation value, the lower the individual credit score.

Through the above a series of steps that will get the four indexes of system clustering, through the AHP - EWM fusion algorithm, determines the relative weight, and standardized data on the results of clustering index of user data, and USES the topsis, determines the optimal vector and vector, the worst in each index are calculated separately, and the Euclidean distance, finally obtained and the optimal value of the recent distance.

Based on the given data and the established scoring model, the distribution of credit scores of these 1000 customers is calculated. The results are shown in the figure below, showing normal distribution as a whole:

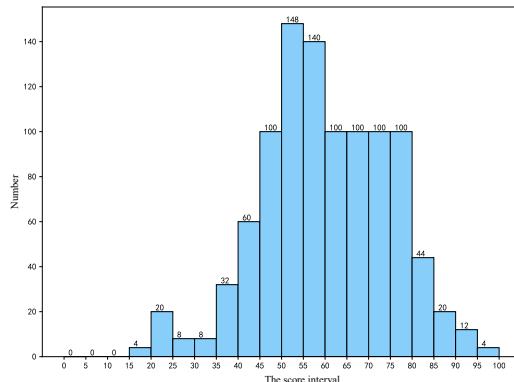


FIG. 2 Histogram of credit score distribution

### 2.4 CLASSIFICATION OF CUSTOMER CREDIT RATING BY K-MEANS CLUSTERING

After obtaining customers' credit scores, we also need to

divide their credit grades. K-means algorithm is used to cluster customers according to the calculated credit scores, and corresponding grades are divided according to the clustering results. The specific process is as follows:

- (1) Randomly select K objects, and each object initially represents the center of a cluster;
  - (2) For each remaining object, assign it to the nearest cluster according to its distance to each cluster center;
  - (3) Recalculate the average value of each cluster and update it to the new cluster center;
  - (4) Repeat 2 and 3 until the criterion function converges.
- By setting different initial values and making several attempts, the optimal results are selected for analysis. The plan of the optimal results is shown in the figure below:

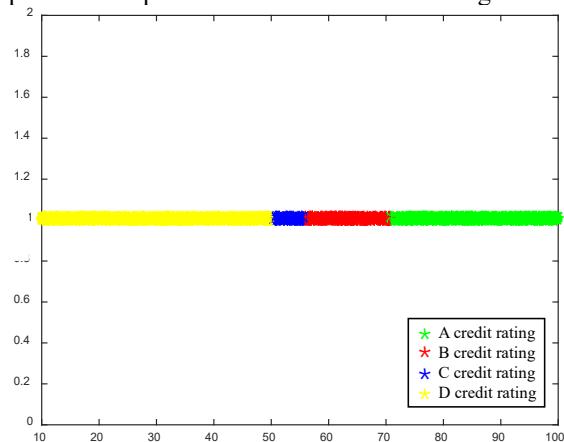


FIG. 3 Clustering results

According to the clustering results, the credit rating of customers can be divided into 4 categories. Here, ABCD is used to represent the four grades. A represents very good credit, B represents good credit, C represents average credit, and D represents poor credit.

Table 11 Level correspondence

credit rating	credit score
A	71.5-100.0
B	57.6-71.5
C	50.2-57.6
D	10-50.2

### 3. TEST THE MODEL

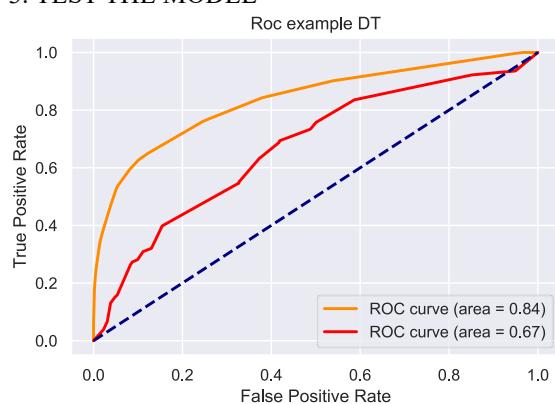


FIG. 3 Roc curve

In the above, in order to verify the generalization of the decision tree model, the KS value is introduced as the basis of the evaluation model. In order to further test the effect of 00 model, Roc curve was introduced. First, 5-fold cross-validation was carried out for the training set data,

and Roc curves were drawn respectively. The average Roc curve was calculated by interpolation, and the results were obtained, as shown in the figure below:

As can be seen from the figure above, the roc curve has a good customer effect of 0.84, while the ROC curve has a bad customer effect of 0.67, which is very good for the credit evaluation model.

## 4. EVALUATION AND PROMOTION OF THE MODEL

### 4.1 ADVANTAGES OF THE MODEL

4.1.1. Data is preprocessed before the evaluation model is established. In particular, one-Hot coding is used for the standardization of disordered discrete data, which effectively improves the data quality and makes the model more accurate.

4.1.2. When constructing the personal credit risk assessment model, a variety of methods are compared, and the current optimal model continues to be optimized with full consideration of a certain principle, and the model results are highly usable.

4.1.3. The personal credit risk assessment model based on AHP-EWM not only considers the subjective experience of experts, but also takes into account the actual situation when determining the index weight, so as to obtain a more objective and reasonable comprehensive weight, avoiding the influence of personal subjectivity.

### 4.2 SHORTCOMINGS OF THE MODEL

Due to the limited time and energy, the established personal credit risk evaluation system also has some shortcomings. The limited factors considered by the model affect the accuracy of the model to a certain extent.

### 4.3 PROMOTION OF THE MODEL

Individual credit decisions can be extended to enterprises, and the research method in this paper can be referred to in the risk assessment of enterprise loans.

## REFERENCE

- [1] Song Y. Research on personal credit evaluation based on Logistic regression and Probit regression [D]. Guangxi Normal University, 2019.
- [2] Liu Junjie. Research on Personal Credit Evaluation based on Big Data [D]. Hainan University, 2020.
- [3] Yang DEjie. Research on personal credit risk Assessment based on deep learning [D]. Central University of Finance and Economics, 2019.
- [4] Jing Y Y. Research on prevention of personal consumption credit risk of China Construction Bank [D]. Shihezi University, 2019.
- [5] Zhao Yun. Personal credit risk assessment based on big data platform [D]. Nanchang University, 2019.
- [6] Li Yimin. Research on optimization strategy of personal credit risk assessment of Binjiang Branch of Z Bank [D]. Zhejiang Gongshang University, 2019.
- [7] Ling FEI. Design and implementation of individual credit decision model based on integrated learning [D]. Zhengzhou University, 2019.
- [8] Guo Chuntao. Research on Personal Credit Evaluation based on Portfolio Model [D]. Tianjin University of Commerce, 2019.
- [9] Tang J Q. Research on default risk measurement of P2P network loan borrowers based on decision tree algorithm

- [D]. Hunan Normal University, 2016.
- [10] Shuai L. Extended research on theory and method of personal credit risk assessment [D]. University of Electronic Science and Technology of China, 2015.
- [11] Miao Jiaming. Personal credit Risk Assessment Model of Commercial Banks based on Data Mining technology and its application [D]. Nanjing University of Finance and Economics, 2015.
- [12] Wang TAO. Research on personal housing credit default risk assessment based on behavior perspective [D]. Ocean University of China, 2013.
- [13] Wu QIlong. Research and application of decision tree algorithm in bank personal credit user classification [D]. Nanjing University, 2012.
- [14] Wang Hao. Research on credit risk Assessment of Individul consumer credit in China's Commercial Banks [D]. Changchun University of Science and Technology, 2011.
- [15] Cheng Laughs. Research on Personal Consumption Credit Risk Evaluation of Commercial Banks in China [D]. South China University of Technology, 2010.
- [16] Fahim Ahmed. K and starting means for k-means algorithm[J]. Journal of Computational Science, 2021, 55.
- [17] LIU Da-zhong, YANG Fei-fei, LIU Sheng-ping. Estimating wheat fractional vegetation cover using a density peak k-means algorithm based on hyperspectral image data[J]. Journal of Integrative Agriculture, 2021, 20(11).
- [18] Geng Duyan, Qin Zhaoxu, Wang Jiaxing, Gao Zeyu, Zhao Ning. Personalized recognition of wake/sleep state based on the combined shapelets and K-means algorithm[J]. Biomedical Signal Processing and Control, 2022, 71(PA).
- [19] Prezelj Jurij, Murovec Jure, Huemer-Kals Severin, Häsler Karl, Fischer Peter. Identification of different manifestations of nonlinear stick-slip phenomena during creep groan braking noise by using the unsupervised learning algorithms k-means and self-organizing map[J]. Mechanical Systems and Signal Processing, 2022, 166.

# Damping Attenuation Characteristics of Valve-Controlled Damping Adjustable Shock Absorber

MeiXiang Zhai<sup>1</sup>, KeMing Chen<sup>1</sup>, Hao Lin<sup>1</sup>, XiaoJun Men<sup>2,\*</sup>

<sup>1</sup> School of Chemical Engineering, North China University of Science and Technology, Tangshan, Hebei, China;

<sup>2</sup> School of science, North China University of Science and Technology, Tangshan, Hebei, China

\*Corresponding Author.

**Abstract:** The main objective of this paper is to study the damping attenuation characteristics of valve-controlled damping adjustable shock absorber. Considering the damping heat generation and dissipation process, the thermodynamic equilibrium model of the shock absorber was established in terms of energy conversion. The damping curves of the shock absorber sample at four different impact velocities were obtained through experiments. Compared with the simulation, it is found that the difference value of the thermal decay rate hits its peak, which is mainly caused by the mechanical friction at low speed. Simultaneously, the experimental results demonstrate that the damping power drops by 25.64% after reaching the thermal equilibrium. Nevertheless, it can still provide damping power of about 281.34J during the one work process. Finally, the study indicates that the shock absorber still has good damping characteristics after thermal decay.

**Keywords:** Adjustable Damping; Shock Absorber; Thermodynamic; Damping Power; Damping Attenuation

## 1. INTRODUCTION

It is universally acknowledged that shock absorber is a fundamental component of automotive suspension to buffer and eliminate road vibration, which plays a significant role in the overall vehicle stability, passenger comfort and driving safety. The working quality of shock absorber itself, i.e. the damping characteristics, has prominent effects on the performance of the suspension. Recently, investigation on damping characteristics of shock absorbers has become the focus at home and abroad. Pengfei Liu et al. [1] proposed a novel resistance switching method to control an electromagnetic damper (EMD) system of a vehicle seat suspension, and researched torque response characteristics of the EMD system in detail. The results demonstrate that the EMD seat was able to reduce the vibration of driver body. Rahmani Mohsen et al. [2] used nonlinear time domain simulations and linear stability analysis to obtain the required equivalent stiffness and damping to be supplied by the shimmy damper. Aydin Ersin et al. [3] investigated earthquake behavior of the structure corresponding to these optimum designs using EI Centro Earthquake (NS) record. As a result, it was shown that the numerical examples that the optimum damper design could be changed based on the vibrations of the designer's constraints. Yan Yong et al. [4] mainly analyzed the effect of the yaw damper with its series stiffness value on the stability and bifurcation type of the railway bogie. Finally, numerical analysis of

corresponding bifurcation diagrams is given to verify the accuracy of the conclusion. Demić M D et al. [5] analyzed the thermal load effects of the shock absorber from the perspective of the damping degradation mechanism, and calculated the dynamic deflection and mechanical work capacity of the suspension under random vibration by simulation. The results demonstrate that the heat generation of the shock absorber decreases with the increase of the vehicle's velocity and the amplitude, and the thermal load effect decreases as the use time increases. Deng Xiaoxu et al. [6] obtained the thermal balance temperature under the parameters of excitation speed, excitation amplitude and quality through a huge influx of experiments, and determined the influence of these parameters on thermal equilibrium. Most specifically, he confirmed the relationship between the parameters by linear regression analysis. However, it's theoretically derived that the damping performance of the shock absorber is affected by various parameters. H. T. Liu et al. [7] established a damping characteristic model of a single-tube inflatable shock absorber, investigated the effects of chamber pressure, oil density and viscosity on damping characteristics by simulation. The results suggest that as the density and viscosity of the oil increases, the damping effect of the shock absorber increases. Ahmad Syuhri et al. [8] studied the damping nature, harvested power and efficiency. The results indicate that the electrical force contributes a lot in low range of amplitudes and frequency. Gatti Gianluca et al. [9] investigated the effect of the nonlinearity and mass ratio of the attachment for both softening and hardening stiffness characteristics using an analytical formulation. Jarrah Majid et al. [10] established a nonlinear FE model to conduct a parametric study on the effect of components geometry on the apparent mechanical properties of the piston metallic damper. It indicates that PMD exhibits stable hysteretic behavior and meet code requirements for low-cycle fatigue. Mikhailov V.P. et al. [11] described a dynamic model of the starting current of an active damper based on a magnetorheological elastomer and derived experimental data for the damper. Dziedziech K. et al. [12] focused on the tuned liquid column damper, both open and sealed, and the identification of its time-varying properties of nonlinear damping, frequency and air pressure identification. Results indicate that the damping ratio is nonlinear, time-varying and depends on the level of vibration. Walsh Kenneth K. et al. [13] performed results of numerical simulations for a five-story base-isolated building subject to four benchmark earthquakes. It

indicates that the RPSD (resetting semi-active stiffness damper) is capable of a similar control performance as the RSASD.

Normally, the previous work utilized the gradient of the damping force peak to evaluate the thermal attenuation characteristics of the actual shock absorber. Actually the real gradient of the damping force peak does not fully represent the thermal attenuation characteristics. The damping characteristic exhibited by the shock absorber diagram is the comprehensive performance of the internal damping characteristics of each damping valve. Therefore, the area enclosed by the dynamometer can be used as the damping force during the unit working period, which is heating power during the unit period.

This paper will take the valve-controlled adjustable damping shock absorber as the research objective. The thermodynamic equilibrium model of the shock absorber will be established in terms of energy conversion to study the characteristics of indicator attenuation before and after thermal equilibrium at four impact velocities.

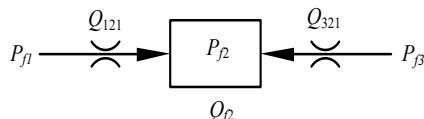
## 2. ESTABLISHMENT AND SIMULATION ANALYSIS OF THERMAL EQUILIBRIUM MODEL

### 2.1. REBOUND STROKE

Considering that the damping force of the shock absorber is significantly larger in the fourth level than the other levels, and the corresponding damping power is also the largest. Thus, this paper will only study the thermodynamic process in the 4th level:

#### (1) Prior to opening the valve

Figure 1 shows the oil passage before opening the valve in the shock absorber, and, the damping force is generated by the fixed orifice corresponding to the rebound valve system.



**Figure 1** The oil passage of rebound stroke prior to opening valve

Taking the piston as the research object, the damping force of the rebound stroke before opening the valve is:

$$F_{f,q} = \left[ \frac{\pi(d_p^2 - d_r^2)}{4a_1} v_{f1} \right]^2 A_p - \left[ \frac{P_{40}V_{40}}{V_{40} + A_c x_{f2}} \left( \frac{\pi d_r^2}{4a_2} v_{f1} \right)^2 + \left[ \frac{\pi(d_p^2 - d_r^2)}{4a_1} v_{f1} \right]^2 \right] A_r$$

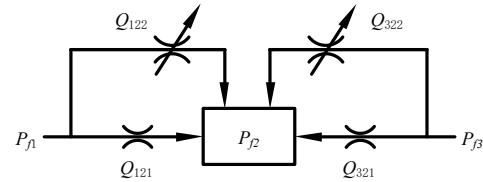
where,  $d_p$  and  $d_r$  denote the diameter of the shock absorber piston and piston rod respectively;  $v_{f1}$  denotes the velocity of the piston during rebound stroke;  $P_{40}$  and  $P_{44}$  respectively denote pre-charge and dynamic pressure of the air chamber;  $V_{40}$  and  $V_{44}$  denote initial and dynamic volume of the chamber respectively;  $A_c$ ,  $A_p$  and  $A_r$  denote the cross sectional area of floating piston, piston and piston rod respectively;  $x_{f2}$  denotes the displacement of floating piston.

$$a_1 = C_v A_{fg} \sqrt{2/\rho}, a_2 = C_v A_{bg} \sqrt{2/\rho}$$

where,  $\rho$  denotes the oil density;  $C_v$  denotes the flow coefficient;  $A_{fg}$  and  $A_{bg}$  denote the fixed orifice area of rebound and compensator valve respectively.

#### (2) After opening the valve

Figure 2 depicts the oil passage after opening the valve,



**Figure 2** The oil passage of rebound stroke after opening valve  
and the damping force can be obtained by the force analysis of the piston:

$$F_{f,h} = \frac{P_{40}V_{40}}{V_{40} + A_c x_{f2}} - \frac{Q_{f3}}{a_4} - \frac{a_2 \left( a_2 - \sqrt{a_2^2 + 4a_4 Q_{f3}} \right)}{2a_4^2} + \frac{Q_{f1}}{a_3} + \frac{a_1 \left( a_1 - \sqrt{a_1^2 + 4a_3 Q_{f1}} \right)}{2a_3^2}$$

where,  $Q_{f1}$  and  $Q_{f3}$  represent the flow through the rebound valve and the compensation valve respectively;  $\mu$  represents dynamic viscosity of oil;

$$a_3 = \frac{\pi \delta_f^3}{6\mu \ln(r_{f2}/r_{f1})}, a_4 = \frac{\pi \delta_b^3}{6\mu \ln(r_{b2}/r_{b1})}$$

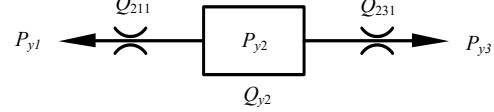
where,  $\delta_f$  and  $\delta_b$  respectively represent the bending deformation of the rebound and the compensation valve discs;  $r_{f1}$  and  $r_{f2}$  and  $r_{b1}$  and  $r_{b2}$  represent the inner and outer radius of the rebound and compensation valve respectively;  $\mu$  denotes the oil dynamic viscosity.

### 2.2 COMPRESSION STROKE

Similar to the analysis process of the rebound stroke, the oil in the compression chamber is mainly divided into two ways during compression stroke: on one hand, the oil flows through the flow valve system into the rebound chamber; on the other hand, the oil flows through the compression valve system into the compensation chamber.

#### (1) Prior to opening the valve

Before opening the valve, the corresponding hydraulic oil circuit is shown in Figure 3,



**Figure 3** The oil passage of compression stroke prior to opening valve

and the damping force can also be obtained:

$$F_{y,kq} = \left[ \frac{\pi(a_p^2 - a_r^2)}{4b_1} v_{y1} \right]^2 A_p + \left[ \frac{P_{40}V_{40}}{V_{40} - A_c x_{y2}} + \left( \frac{\pi a_p^2}{4b_2} v_{y1} \right)^2 - \left( \frac{\pi(a_p^2 - a_r^2)}{4b_1} v_{y1} \right)^2 \right] A_r$$

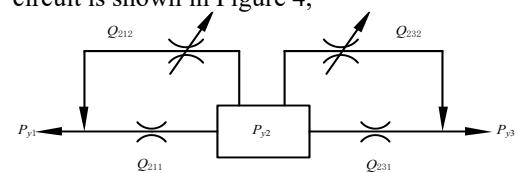
where,  $v_{y1}$  is the velocity of the piston rod during compression stroke;  $x_{y2}$  is the displacement of floating piston during compression stroke.

$$b_1 = C_v A_{lg} \sqrt{2/\rho}, b_2 = C_v A_{yg} \sqrt{2/\rho}$$

where,  $A_{lg}$  and  $A_{yg}$  are the fixed orifice area of flow-off and compression valve respectively.

#### (2) After opening the valve

After opening the valve, the corresponding hydraulic oil circuit is shown in Figure 4,



**Figure 4** The oil passage of compression stroke after opening valve  
and the damping force can be obtained:

$$F_{sl} = \left[ \frac{Q_{y1}}{b_3} + \frac{b_1(b_1 - \sqrt{b_1^2 + 4b_3 Q_{y1}})}{2b_3^2} \right]^2 A_p + \left[ \frac{P_{sl} V_{40}}{V_{40} - A_p x_{y2}} + \frac{Q_{y3}}{b_4} + \frac{b_2(b_2 - \sqrt{b_2^2 + 4b_4 Q_{y3}})}{2b_4^2} \right] A_s - \frac{Q_{y1}}{b_3} - \frac{b_1(b_1 - \sqrt{b_1^2 + 4b_3 Q_{y1}})}{2b_3^2} A_s$$

where,  $Q_{y1}$  and  $Q_{y3}$  represent the flow through the flow valve and compression valve respectively;

$$b_3 = \frac{\pi \delta_l^3}{6\mu \ln(r_{l2}/r_{l1})}, b_4 = \frac{\pi \delta_y^3}{6\mu \ln(r_{y2}/r_{y1})}$$

where,  $\delta_l$  and  $\delta_y$  denote the bending deformation of the flow valve and compensation valve discs;  $r_{l1}$ ,  $r_{y2}$ ,  $r_{l1}$  and  $r_{l2}$  denote the inner and outer radius of the compression and flow valve.

The value of work done by the shock absorber during the working cycle, i.e., the heating work during per unit cycle is given by:

$$W_z = FS$$

where,  $F$  denotes the damping force;  $S$  denotes the displacement of the piston rod.

The thermal energy generated by the damping force will be shifted to two parts: one part is absorbed by the shock absorber body and thereby causing its own temperature to rise; furthermore, the other part is diffused into the atmosphere by heat exchange. In accordance with the conversion of energy, the thermal equilibrium equation of the shock absorber can be obtained:

$$W_z = Q_{ws} + Q_{sr}$$

where,  $W_z$  is the damping work;  $Q_{ws}$  is the work causing the temperature rise;  $Q_{sr}$  is the work of external heat dissipation

Wherein, the temperature rise equation of the shock absorber body is:

$$Q_{ws} = (C_l M_l + C_s M_s) \frac{dT_{oil}}{dt}$$

where,  $C_l$  and  $C_s$  denote the specific heat of oil and cylinder respectively;  $M_l$  and  $M_s$  denote the mass of oil and cylinder respectively;  $T_{oil}$  denotes the oil temperature.

### 2.3 HEAT CONDUCTION MODEL

During the heat dissipation process of the shock absorber, the heat in the working cylinder is transmitted to the housing of the reservoir cylinder by means of heat conduction. Therefore, the heat flux of the heat conduction process yields:

$$\Phi_{cd} = \frac{T_{n1} - T_{w2}}{R_{cd}}$$

where,  $T_{n1}$  is the temperature of the inner wall of the working cylinder;  $T_{w2}$  is the temperature of the outer wall of the reservoir cylinder.

The thermal resistance of the  $R_{cd}$  in the Eq.(12), and the corresponding calculation formula thereof can be obtained:

$$R_{cd} = \frac{1}{2\pi\lambda_g l_n} \ln \frac{r_{n1}}{r_{n2}} + \frac{1}{2\pi\lambda_g l_w} \ln \frac{r_{w1}}{r_{w2}} + \frac{1}{2\pi\lambda_y l_{cy}} \ln \frac{r_{w1}}{r_{n2}}$$

where,  $r_{n1}$  and  $r_{n2}$ , denote the inner and outer radius of the work cylinder respectively;  $r_{w1}$  and  $r_{w2}$  denote

the inner and outer radius of reservoir cylinder respectively;  $\lambda_g$  and  $\lambda_y$  denote the cylinder and oil thermal conductivity respectively;  $l_n$ ,  $l_w$  and  $l_{cy}$  denote the effective length of work cylinder, reservoir cylinder and overlapping chambers.

### 2.4 HEAT CONVECTION MODEL

The heat transfer process of the oil in the working chamber is forced convection heat transfer in the tube, and the corresponding heat transfer model can be expressed as:

$$\Phi_{dl} = h_{dl} A_n (T_y - T_{n1})$$

where,  $h_{dl}$  denotes the coefficient of forced convection heat transfer;  $A_n$  denotes the heat exchange area between the inner wall of the working cylinder and the oil;  $T_y$  denotes the oil temperature;  $T_{n1}$  denotes the inner cylinder temperature.

### 2.5 THERMAL RADIATION MODEL

If the surface of the reservoir cylinder is coated with black paint, the heat transfer process of the external radiation is supposed to be considered. In fact, the radiant heat flux of a black object can be obtained by the following formula:

$$\Phi_r = \varepsilon c_b A_{w2} \frac{T_{w2}^4 - T_{air}^4}{100^4}$$

where,  $\varepsilon$  denotes the emissivity of external radiation of reservoir cylinder, here,  $\varepsilon=0.75$ ;  $T_{w2}$  and  $T_{air}$  denote the temperature of the surface of the reservoir cylinder and the external environment respectively;  $c_b$  denotes the emissivity of the reservoir cylinder, here,  $c_b = 5.67$ ;  $A_{w2}$  denotes the radiation area of the reservoir cylinder.

Based upon the thermodynamic analysis above and the series-parallel relationship between the numerous thermal resistances in the heat transfer process of the shock absorber, hence, the expression of the total heat flux during the heat transfer of the shock absorber is as follows:

$$Q_{sr} = \frac{T_y - T_{air}}{R_z}$$

where,  $R_z = R_{dl} + R_{cd} + R_b$ ;  $R_z$  denotes the total resistance during heat transfer;  $R_{dl}$  denotes the thermal resistance of heat convection;  $R_b$  denotes the thermal resistance of heat transfer from the surface of the reservoir cylinder.

## 3.EXPERIMENTAL ANALYSIS OF ATTENUATION CHARACTERISTICS

The whole shock absorber was taken as the experimental object, and the experiment was carried out according to the test procedure stipulated by the national standard QCT545-1999. First, the shock absorber sample is placed in a normal temperature environment for more than 6 hours. Secondly, according to the performance of the shock absorber, the stroke of the piston rod is  $\pm 37.5$  mm, and the excitation frequency of the actuator is set to 0.556 Hz, 1.12 Hz, 1.668 Hz and 2.207 Hz. The specific process of the experiment is as follows:

- (1)Fixing the lower end of the shock absorber on the servo cylinder of the test bench, so that the upper end of the piston rod is suspended;
- (2)Start the hydraulic power source to supply oil to the test bench;
- (3)Input the control command on the host computer

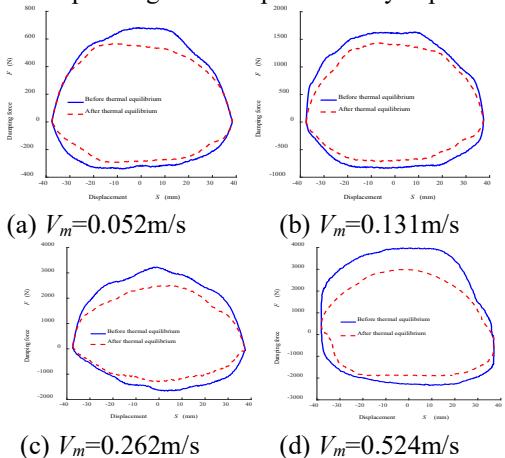
interface, find the middle position by measuring the highest position and the lowest position of the piston rod, and initialize the position of the piston rod;

(4) Fixing the upper end of the piston rod to the force sensor;

(5) Input the corresponding sinusoidal excitation signal on the upper monitor and set the experiment value every 2 minutes;

(6) Export the test results and perform data processing to output the damping force and obtain temperature rise curve.

The electro-hydraulic servo test bench is utilized to collect three signals of the real-time temperature, damping force and piston rod displacement of the shock absorber sample. Moreover, in accordance with the test results, as are shown in Figure 5, the change of the indicator diagram corresponding to the impact velocity is plotted.



**Figure 5** Damping experimental values before and after thermal equilibrium

It can be seen from Figure 5 that although the

**Table 1** The comparison of damping energy before and after thermal equilibrium

Velocity (m/s)	Peak value of damping force (/N) and damping work at 25°C (/J)			Peak value of damping force (/N) and damping work (/J) after thermal equilibrium			Thermal decay rate	
	Rebound	Compression	Damping work	Rebound	Compression	Damping work	Experiment (%)	Simulation (%)
0.052	699.79	-347.72	59.95	567.29	-332.04	52.48	12.46	9.02
0.131	1631	-845.08	143.45	1441.6	-717.52	123.77	13.72	15.52
0.262	3121.6	-1675.5	264.39	2509.0	-1306.10	209.23	20.86	21.05
0.524	3982.7	-2218.0	378.35	2991.8	-1779.2	281.34	25.64	27.67

**Table 2** The comparison of experimental and simulated values of damping force peak and damping work

Velocity(m/s)	Peak value difference of damping force (/N) and work difference (/J) at 25 °C			Peak value difference of damping force (/N) and work difference (/J) after thermal equilibrium		
	Rebound	Compression	Damping work	Rebound	Compression	Damping work
0.052	35.09	15.9	1.29	-37.47	30.04	-0.89
0.131	30.7	15.58	-3.2	37.9	16.79	-0.11
0.262	93.8	74.2	-24.98	23.9	41.7	-19.25
0.524	145	302.2	-6.55	224.3	93.6	2.94

#### 4.CONCLUSIONS

In this paper, the damping heat generation and dissipation process of the shock absorber were considered comprehensively. Then, the thermodynamic equilibrium model of the shock absorber was established. Finally, the damping work attenuation characteristics of the shock absorber at four impact speeds were studied. The results

experimental curve of the dynamometer is not as saturated and round as the simulated one, there is no sustained zero damping, and the damping changes smoothly and continuously. Hence, it can be considered that the vibration damping effect is well. Furthermore, the work attenuation in various degrees after thermal equilibrium verifies that of the simulation.

Most specifically, the peak values of the damping force and the work before and after thermal equilibrium are shown in Tables 1 and 2 respectively. Note that the maximum difference between the experimental and the simulated results of the damping force peak is 302.2N, accounting for around 13.6% of the compressive force peak when it is at 0.524m/s and 25°C during the compression stroke. On the other hand, the maximum difference between the experimental and simulated value of the damping work is 24.98J, occupying approximately 9.45% of the work per unit cycle, which occurs at the impact velocity of 0.262m/s prior to the thermal equilibrium. In addition, the experimental results suggest that the thermal decay rate increases with the increase of the impact velocity and thereby verifying the correctness of the simulation model.

By comparison, it can be found that the experimental values of other thermal decay rate except at 0.052m/s are smaller than the simulated ones, but the difference at the same excitation speed does not exceed 5%. Then, the analysis shows that the difference is mainly owing to the work caused by mechanical friction at low speed. Furthermore, the damping power drops by 25.64% after achieving thermal equilibrium at the impact velocity of 0.524m/s, however, it can still provide the damping power of 281.34J during one work process. That is, the experiment demonstrates that the shock absorber retains good damping characteristics after thermal decay.

demonstrate that the higher the impact velocity, the larger the damping power, and the greater the damping power attenuation after reaching the thermal equilibrium. The experimental results also indicate that the thermal decay rate increases with the increase of the impact velocity, which verifies the correctness of the simulation model.

**REFERENCES**

- [1] Pengfei Liu, Xiangjun Xia, and Nong Zhang. (2019), “Torque response characteristics of a controllable electromagnetic damper for seat suspension vibration control”, *Mechanical Systems and Signal Processing*, Vol.133 No.1, pp.106-120.
- [2] Rahmani Mohsen and Behdinan Kamran. (2019), “On the effectiveness of shimmy dampers in stabilizing nose landing gears”, *Aerospace Science & Technology*, Vol.91 No.1, pp.272-286.
- [3] Aydin Ersin, Öztürk Baki, and Dutkiewicz Maciej. (2019), “Analysis of efficiency of passive dampers in multistorey buildings”, *Journal of Sound & Vibration*, Vol.439 No.1, pp.17-28.
- [4] Yan Yong, Zeng Jing and Huang Caihong. (2019), “Bifurcation analysis of railway bogie with yaw damper”, *Archive of Applied Mechanics*, Vol.89 No.7, pp.1185-1199.
- [5] Demić, M. D. and Diligenski, Đ. M. (2016), “Numerical simulation of shock absorbers heat load for semi-active vehicle suspension system”, *Thermal Science*, Vol. 20 No.5, pp.1725-1739.
- [6] Xiaoxu, Deng. (2017), “Simulation and experimental study on the thermal balance of twin tube shock absorber”, *Jilin University*, pp. 20-51.
- [7] Hongtuo, Liu., Fangwei, Xie., and Kai, Zhang. (2018), “Effect of air chamber and oil properties on damping characteristics of single-tube pneumatic shock absorber”, *International Journal of Structural Integrity*, Vol. 9 No.1, pp.27-37.
- [8] Ahmad Syuhri, Widyono Hadi and Skriptyan N. H. Syuhri. (2018), “Damping properties and energy evaluation of a regenerative shock absorber”, *International Journal on Interactive Design and Manufacturing (IJIDeM)*, Vol.12 No.4, pp.1385-1397.
- [9] Gatti Gianluca. (2018), “Fundamental insight on the performance of a nonlinear tuned mass damper”, *Meccanica*, Vol.53 No.1-2, pp.111-123.
- [10] Jarrah Majid, Khezrzadeh Hamed and Mofid Massood. (2019), “Experimental and numerical evaluation of piston metallic damper (PMD)”, *Journal of Constructional Steel Research*, Vol.154 No.1, pp.99-109.
- [11] Mikhailov V.P., Bazinenkov A.M. and Dolinin P.A. (2018), “Dynamic Modeling of an Active Damper”, *Russian Engineering Research*, Vol.38 No.6, pp.434-437.
- [12] Dziedziech K., Ghosh A. and Iwaniec J. (2018), “Analysis of tuned liquid column damper nonlinearities”, *Engineering Structures*, Vol.171 No.1, pp.1027-1033.
- [13] Walsh Kenneth K., Sallar Grace and Steinberg Eric P. (2017), “Modeling and Validation of a Passive Resettable Stiffness Damper”, Vol.143 No.2, pp.733-758.

# Research On Dynamic Programming Model of Ordering and Transportation of Raw Materials Based on Production Enterprises

Xiaoxiao Dong<sup>1,2</sup>, Haowei Yan<sup>1,3</sup>, Liya Chen<sup>4</sup>

<sup>1</sup>Laboratory of Engineering Calculation and Simulation Innovation, North China University of Technology, Tangshan 063210, China;

<sup>2</sup>College of Science, North China University of Technology, Tangshan 063210, China;

<sup>3</sup>School of Architectural Engineering, North China University of Technology, Tangshan 063210, China;

<sup>4</sup>School of Chemical Engineering, North China University of Technology, Tangshan 063210, China

\*Corresponding Author.

**Abstract:** For a long time, the ordering and transportation of goods is always the key problem in the production process of enterprises. How to reduce the cost of capital investment and increase revenue in the case of decision-making is often a major concern of enterprises. Choosing different suppliers and transporters, raw material procurement costs and transshipment losses will be different. In order to reduce production costs, some enterprises also increase labor productivity; or the production efficiency of enterprises remains unchanged, the way to reduce unit consumption to increase profits. The discrete stochastic dynamic programming model is used to find the optimal solution of the scheme step by step recursively based on the objective function and boundary equation. Since the final results obtained by the two algorithms are the same, for example, if the dynamic programming model only uses the inverse recursive method for solving, the solving speed will be greatly reduced when there are many stages. From the perspective of practical production, the bi-directional recursive method is used to solve the model, which can speed up the solving rate of multi-stage dynamic programming model.

**Keyword:** Fuzzy Comprehensive Evaluation Model; Discrete Dynamic Programming Model; Reverse Recursive Method; Simulation

## 1. INTRODUCTION

In the face of raw materials and enterprise production problems, often multi-process optimization problem. Suppliers and transporters have corresponding constraints, how to control variables through objective function to maximize benefits is the key. The selection of raw materials is divided into two stages. One is the selection of suppliers, and the other is the selection of transporters. Because of the discreteness and randomness between the two choices, the dynamic programming model is a method to solve the optimization of multi-stage decision-making process, which transforms the multi-stage problem into a series of interrelated single-stage problems and then solves them one by one. Therefore, this paper establishes a discrete stochastic dynamic programming model to obtain the maximum profit.

The key point of dynamic programming model is to transform a big problem into a series of sub-problems and

solve them one by one. In this multi-stage decision process, each stage selection only depends on the current state to determine the output state. After the decision of each stage is determined, a decision sequence is formed, which determines the final solution of the problem, and the overall optimality is derived by the optimal progression of the subproblem. However, in the sub-process, the selection of suppliers will also affect the selection of transshippers, so it is necessary to establish the transfer relationship of the state before and after the stage, namely the state transfer equation.

In the solution of dynamic programming, the sequential recursive method or the inverse recursive method are often used to calculate the optimal solution of each sub-process, and the final results of the sequential recursive method and the inverse recursive method are exactly the same. Since there are only two sub-processes in this problem, the calculation speed will not be too slow if only one of them is used. But considering the real life is often multi-process time, the solution speed will be greatly reduced. Therefore, on this basis, the use of two-way recursive method to speed up the problem solving speed, so that the model is more widely used.

## 2. DISCRETE DYNAMIC PROGRAMMING MODEL FOR ORDERING AND TRANSPORTATION OF RAW MATERIALS

### 2.1 DETERMINATION OF STAGE FUNCTIONS

Considering the progress sequence of product manufacturing, the supply, transportation and production of raw materials can be naturally divided into three stages, so that the problem can be solved and optimized in the order of stages.[1]

Assuming suppliers supply A, B, C three different raw materials. The enterprise choose suppliers as stage one, expressed by S1. A certain transshipment company with low transshipment loss rate meeting the production demand of the enterprise is taken as stage II, which is represented by S2. Enterprises rely on suppliers and transporters to provide sufficient raw materials and advanced production equipment to manufacture products for stage three, using S3 to represent the variables in each stage can be expressed as  $k = 1, 2, \dots, n$ . (Here,  $n = 3$ ) represents a total of three stages.[2] As shown in Fig. 1.

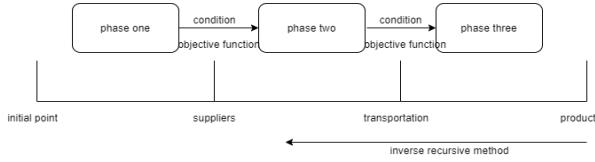


Figure 1. Stage simplified flow chart

According to the requirements of raw materials and loss rate in the actual production of enterprises, the corresponding index function is determined, that is, the supplier supplies the raw materials to one of its corresponding transporters efficiently. When the loss rate of the transporter is small, the raw materials are transported to the enterprise. The capacity of the product manufactured by the enterprise is expressed by  $V_{k,n}(x_k, u_k, x_{k+1}, \dots, x_{n+1})$ ,  $k = 1, 2, \dots, n$ . Since the dynamic programming model is discrete,  $V_{k,n}$  can be expressed as functions about  $x_k, u_k, V_{k+1,n}$ , denoted as:

$$\begin{aligned} V_{k,n}(x_k, u_k, x_{k+1}, \dots, x_{n+1}) \\ = \partial_k(x_k, u_k, V_{k+1,n}(x_{k+1}, u_{k+1}, \dots, x_{n+1})) \end{aligned}$$

In this formula, the function  $\partial_k$  is monotonic for the index variable  $V_{k+1,n}$ .

After quantitative analysis on the supply and transfer of raw materials by suppliers and transporters, it is not difficult to find that indicators in each stage can be considered comprehensively. In the actual process, the stage index  $s$  of stage S depends on the state  $x_s$  and the decision  $u_s$ , which is expressed by  $v_s(x_s, u_s)$ . The index function is composed of the following:

$$V_{k,n}(x_k, u_k, x_{k+1}, \dots, x_{n+1}) = \sum_{s=1}^n v_s(x_s, u_s)$$

## 2.2 DYNAMIC PROGRAMMING MODEL OF ORDERING AND TRANSPORTATION BASED ON RAW MATERIALS OF MANUFACTURING ENTERPRISES

After establishing the decision variables in the three stages of raw material supply, transportation and manufacturing, the strategy of a certain stage  $S_k$  is obtained by combining the sequence of each decision variable  $u_k$ . Here, we record the whole process strategy of each stage of raw materials starting from the initial state  $x_1$  in stage one  $S_1$  as follows: The strategy from the first stage to the end stage can be recorded as  $P_{1n}(x_1)$ , namely

$$P_{1n}(x_1) = \{u_1(x_1), u_2(x_2), \dots, u_n(x_n)\}$$

The strategy from the state  $x_k$  of stage k to the second half of the whole process ending state is denoted as:

$$P_{kn}(x_k) = \{u_k(x_k), \dots, u_n(x_n)\}, k = 1, 2, \dots, n - 1$$

The strategy from stage k to stage  $k+1$  is denoted as:

$$P_{k(k+1)}(x_k) = \{u_k(x_k), \dots, u_{k+1}(x_{k+1})\}$$

In different processes of each stage, with the change of state, the optional strategy is within a certain range, and the set of strategies that can be allowed by the reality is defined as the set of permissible strategies, which is represented by  $P_{1n}(x_1), P_{kn}(x_k), P_{k(k+1)}(x_k)$ .

In this model, in order to make the index function reach the optimal result required by the enterprise, the strategy in the whole process is linked to the optimal strategy, denoted as:  $P_{kn}^* = \{u_k^*, \dots, u_n^*\}$

where  $P_{1n}^*$  is the optimal strategy in the whole process.

Starting from the initial state  $x_1$ , the state is updated according to the optimal strategy  $P_{1n}^*$  and the state transition equation. The formed state set is called the optimal trajectory, which is denoted as:

$$\{x_1^*, x_2^*, \dots, x_{n+1}^*\}$$

Since the transformation process of raw materials between stage  $S_k$  and stage  $S_{k+1}$  has been determined, and as the end user, enterprises require to purchase class A raw materials as much as possible and class C raw materials as little as possible in stage 1, and reduce the cost of transportation and storage in stage 2, and require suppliers to minimize the transshipment loss rate of the corresponding transshipment suppliers, while ensuring that the production capacity of their own enterprises is improved, that is, the state  $x_k$  of  $S_k$  in each stage is transferred, and the state  $x_{k+1}$  of the next stage has been determined.

The state transition equation is used to express this evolution law as follows:

$$x_{k+1} = T_k(x_k, u_k), k = 1, 2, \dots, n$$

According to the index function of the state transition equation, it can be expressed as the function of state  $x_k$  and strategy  $P_{kn}$ , namely  $V_{k,n}(x_k, P_{kn})$ . Combining the above known state  $x_k$ , the optimal value function of the index function  $V_{k,n}$  with respect to the policy  $P_{kn}$  is obtained, denoted by  $f_k(x_k)$ , that is:  $f_k(x_k) = \min V_{k,n}(x_k, p_{kn})$

After obtaining the optimal strategy of the whole process, the model needs to be optimized many times to obtain the optimal solution of the procurement and transfer scheme between suppliers, transporters and enterprises. Each optimization solution is a mechanical repetition process with the same principle, which can be regarded as a recursive equation when using computer-aided simplified calculation, namely:

$$\left\{ \begin{array}{l} f_{n+1}(x_{n+1}) = 1 \\ f_k(x_k) = \min \{v_k(x_k, u_k) * f_{k+1}(x_k)\}, k = 1, 2, \dots, n \\ V_{k,n}(x_k, u_k, x_{k+1}, \dots, x_{n+1}) = \sum_{s=1}^n v_s(x_s, u_s) \\ P_{kn}(x_k) = \{u_k(x_k), \dots, u_n(x_n)\}, k = 1, 2, \dots, n - 1 \\ P_{kn}^* = \{u_k^*, \dots, u_n^*\} \\ x_{k+1} = T_k(x_k, u_k), k = 1, 2, \dots, n \\ f_k(x_k) = \min V_{k,n}(x_k, p_{kn}) \\ f_{n+1}(x_{n+1}) = 1 \\ f_k(x_k) = \min \{v_k(x_k, u_k) * f_{k+1}(x_k)\}, k = 1, 2, \dots, n \end{array} \right.$$

According to the optimality principle analysis of dynamic programming, this model conforms to the actual production situation. Formula (5) reflects the pros and cons of supplier and transporter selection scheme, namely the objective function of the model.

## 2.3 MODEL SIMULATION

The inverse order method is used to solve the model: the inverse order recursion is used to solve the model. According to the characteristics of time and space, the complex problems are divided into three interrelated stages. After selecting the direction of the system, the reverse calculation is carried out from the end point to the



Figure 3. Resolving flow chart

### 3.3 BASIC EQUATIONS FOR THE IMPROVED MODEL ALGORITHM:

$$ff_k(x_{k+1}) = \max / \min [v_k(x_{k+1}, u_k) + ff_{k-1}(x_k)], k = 1, 2, \dots$$

$$fb_j(y_j) = \max / \min [v_j(y_j, u_j) + fb_{j+1}(y_{j+1})], j = n, n-1, \dots$$

Starting condition:  $ff_0(x_1) = 0$ ;

Terminal conditions:  $fb_{n+1}(y_{n+1}) = 0$ ;

### 3.4 SIMULATION OF MULTI-DECISION ALGORITHM

Suppose there are  $k$  stages and  $k$  decisions are needed. The stage and stage variables are  $k$  and the state variables are  $x_k$ . Accordingly, the decision variables  $P_k$ , [7] the

optimal strategy  $P_k^n$ , the index function  $V_k$  and the optimal value function  $f_k$  are expressed as follows:

$$P_k(x_k) = |u_k(x_k), u_{k+1}(x_{k+1}), \dots, u_n(x_n)|$$

$$P_k^*(x_k) = |u_k^*(x_k), u_{k+1}^*(x_{k+1}), \dots, u_n^*(x_n)|$$

$$V_k = V_k(x_k, u_k, x_{k+1}, u_{k+1}, \dots, x_n, u_n)$$

$$V_k = v_k(x_k, u_k) \oplus v_{k+1}(x_{k+1}, u_{k+1}) \oplus \dots \oplus v_n(x_n, u_n)$$

$$f_k(x_k) = optV_{k,n}(x_k, u_k, x_{k+1}, u_{k+1}, \dots, x_n, u_n)$$

Since the final results obtained by the reverse recursive and sequential recursive algorithms are the same, the bidirectional recursive method is used to solve the model to speed up the problem solving rate of the multi-stage dynamic programming model.

### 4. CONCLUSION

Based on the dynamic programming model, this paper establishes a discrete random dynamic programming model according to the actual situation of suppliers and transporters. Through the analysis of the supply characteristics of raw material suppliers and transporters, three evaluation indexes are summarized as supply volume, supply rate and procurement cost. By changing the constraint conditions of the scheme, the company's profit is increased. Some enterprises also increase their profits by improving labor productivity or reducing unit consumption, so the raw material ordering scheme needed by enterprises will also change.[8]

Although the focus of the two problems is different, but

the objective function is the same, and the boundary conditions are small, are recursive to find the optimal solution. Here, the two are put into a discrete random dynamic programming model to consider and analyze. According to the dynamic programming model, the optimal solution is calculated by multiple iterations of reverse order recursion, and then the new ordering and transportation schemes that meet the requirements of enterprises are better planned.

From the perspective of actual production, the bidirectional recursive method is used to solve the model, which can speed up the solving rate of the multi-stage dynamic programming model, and more in line with the needs of enterprise production.

### REFERENCES

- [1] Li Yinya. Study on the optimization of strategic petroleum reserve under supply disruption scenario. Jiangsu: Jiangsu University of Science and Technology, 2016.
- [2] Li Jie, Fei Shilong, Liu Zhaopeng. The application of dynamic programming in the shortest path problem of logistics distribution. Journal of Jilin University of Chemical Technology, 2016 (11): 117-121.
- [3] Zhou Xiang. Feasibility study of crude oil blending scheme using scheduling optimization model. Petroleum refining and chemical industry, 2018, 49 (2): 84-89.
- [4] Jiang Qiyuan, Mathematical Model. Second Edition, Beijing: Higher Education Press, 1993, 297-305 pp.
- [5] Liu Qun, inventory management problem analysis and control strategy under supply chain management environment, value engineering, 2009, (5), 69-72.
- [6] Shi Guohong, Yu Chenglong, Gong Wenwei. Research on supply chain coordination based on production uncertainty and demand uncertainty. Industrial engineering and management, 2011, 16 (2): 1-4.
- [7] Luo Bing, Xiong Zhongkai, Yang Xiutai. EOQ model when inventory affects sales rate and theoretical demand is linear time-varying function. China Management Science, 2002, 10 (6): 66-71.
- [8] Wang Shengdong, Zhou Yongwu. EPQ model based on raw material ordering and preventive maintenance interruption. Journal of Systems Engineering, 2005, 20 (4): 381-386, 418.

# The Design and Management Model of Airport Taxi Based on Goal Planning

Zehui Dou<sup>1, 2, 3</sup>, Yumei Gong<sup>1, 2, 5</sup>, Renfang Wang<sup>1, 4</sup>

<sup>1</sup> Mathematical Modeling Innovation Lab, North China University of Science and Technology, Tangshan 063210, China;

<sup>2</sup> Engineering Computing and Simulation Innovation Lab, North China University of Science and Technology, Tangshan 063210, China;

<sup>3</sup> College of Artificial Intelligence, North China University of Science and Technology, Tangshan 063210, China;

<sup>4</sup> School of Science, North China University of Science and Technology, Tangshan 063210, China;

<sup>5</sup> School of materials science and Engineering, North China University of Science and Technology, Tangshan 063210, China

\*Corresponding Author.

**Abstract:** The design of airport taxi involves two aspects, On the one hand, in order to minimize the phenomenon of passengers queuing and boarding, it was decided to distribute the parking spots on both sides of the parallel lanes, set up S-shaped entrances in the queuing area to evacuate passengers, and cross the parallel lanes to ensure the safety of passengers; on the other hand,, In order to improve the efficiency of taxis carrying passengers, the parking area of taxis is set to S type. Based on the setting of this parking lot, taking the number of boarding points, the number of passengers and the number of taxis as decision variables, and the efficiency of the ride as the optimal goal, a constrained ride efficiency model is obtained. It is found that under a certain passenger flow, set 10 There are two parking spaces and five "boarding points", and the efficiency of the ride is at its maximum at this time. Determine the average return per unit time of normal passenger return to the urban area and the average return per unit time of short-distance passenger travel, and establish a model with the smallest difference, that is: the difference between the return of a taxi "short-distance return + return to the urban area" and the return of direct return to the urban area If the value is the smallest, it is found that a short-distance taxi carrying passengers only needs to return within  $2T_0$  of the passenger departure, and then wait for t time to "prioritize" passengers.

**Keywords:** Single objective optimization; Airport taxi; Design and Management

## 1. INTRODUCTION

At present, the national economy as a whole is in a state of growth, and the consumption capacity of citizens has increased, and more people choose to travel by air passenger. Passengers arriving at the airport will choose airport buses, intercity railways, taxis, etc., in order to get to the destination in the city. Among them, taxis are not restricted by time and destination, and it is very convenient without transfers, so this way of travel Favored by flight passengers, according to brief statistics, the daily average daily operation volume of taxis to and from Zhengzhou Airport reaches nearly 2, 000, which is one of the main ways for passengers to collect and distribute.

There are often two parallel lanes in the "ride area" of the airport. It is necessary to set up a "boarding point"

according to the actual situation, and arrange taxis and passengers reasonably under the conditions of ensuring the safety of vehicle passengers, so as to alleviate the situation of taxis and passengers waiting in line. Maximize the total ride rate.

Taxi's passenger income is related to the mileage, and the destination of passengers is far or near. Under the premise of not being able to refuse to carry and select passengers, taxi drivers are allowed to carry passengers multiple times. Therefore, the management department needs to give "priority" to the return of certain short-distance passenger-carrying taxis in order to balance the total revenue of each taxi driver, and establish a reasonable "priority" arrangement plan based on this condition.

## 2. EXPERIMENTAL

### 2.1 SELECTION OF THE BEST "PICK-UP POINT"

From the perspective of airport management, under the premise of ensuring the safety of passengers and vehicles, reasonable ride rules should be set up and "boarding points" should be arranged to improve overall ride efficiency. Therefore, we take ride efficiency as the final optimization goal, and the number of boarding points, the number of passengers arriving at the airport within a unit time, and the number of taxis arriving at the airport within a unit time are used as dependent variables for analysis.

At present, most domestic airport taxi pick-up areas are single-lane parking, matrix multi-lane parking, and oblique parking. [1][2] The single-lane taxi pick-up method causes each stop due to lane restrictions. The number of taxis in the parking area is small, and the efficiency of taxi departure cannot be guaranteed; the multi-lane taxi pick-up method can accommodate multiple taxis at one time, which is convenient for passengers to carry luggage, and to a certain extent avoids the mixed traffic situation. The safety is high; the oblique taxi pick-up method can effectively improve the efficiency of the ride and effectively reduce the interference between leaving vehicles. Since there are only two parallel lanes in the airport "ride area" given in the question, a matrix-type lane parking and pick-up system is adopted.

First of all, in order to improve the convenience and efficiency of the ride, we set the "boarding point" on both sides of the road to reduce the time wasted due to the front and rear obstruction of the taxi, and double the ride

efficiency of the "riding area" [3]. Secondly, in order to ensure the safety of vehicles and passengers and avoid congestion caused by pedestrians crossing the road, a pedestrian bridge connecting the two sides of the parallel road can be built above the two lanes, and the position of the overpass is right at the airport exit station, avoiding the opposite side" The "boarding point" generates unnecessary walking distance due to the long distance, which improves the convenience and overall riding efficiency of passengers on the opposite side, while avoiding the low traffic of the "boarding point" on the opposite side. If there are many "boarding points" on one side of the road, it will cause the vehicle with passengers in the rear to be blocked by the vehicle in front, which will reduce the efficiency of the ride. Therefore, considering the riding situation, set up multiple "boarding points" on one side. k, the specific idea is shown in Figure 1.

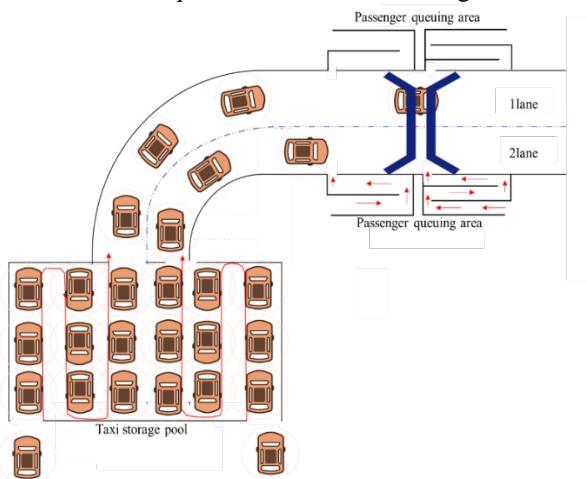


Figure 1 Schematic diagram of airport taxi passenger arrangement 1

This operation method of picking up passengers operates as follows: Arrange a number of taxi berths in the storage tank. At a certain moment, the gates are opened, and the corresponding number of taxis enter the pick-up area through the parallel lanes and stop at the corresponding parking spots. After the taxi stops, the outbound passengers will enter the pick-up area one by one. After passengers board the car, the two lanes of taxis will leave in turn, until the last taxi leaves the parking area, the gates will be opened again, and the pick-up area will enter the taxi.

## 2.2 AIRPORT OPERATION EFFICIENCY MODEL BASED ON SINGLE OBJECTIVE OPTIMIZATION

### 1. The establishment of the model

In order to improve the operating efficiency of the airport, it is necessary to consider both passengers and taxis, while reducing the waiting time for passengers and taxis at the airport.

**Step1:** Determination of optimization goals. The waiting time of a taxi driver is defined as the time between the taxi entering the station and the time when the passenger is loaded, and the waiting time of the passenger is defined as the sum of the waiting time for boarding and the boarding time. Assuming that during the operation of the airport, passengers carry taxis at a stable rate as a whole, set the

riding efficiency as  $\eta$ , the total number of passengers as a, and the total number of taxis as b, then:

$$\max \eta = \frac{1}{\sum_{i=1}^a t_i/a + \sum_{j=1}^b T_j/b} \times \frac{2k}{\alpha^k} \quad (1)$$

Because passengers on both sides get on the bus at the same time for two "boarding points" when taking a taxi, the riding efficiency is increased by four times. Among them,  $t_i$  is the slower boarding speed of the two "boarding points" on one side. The total waiting time of passengers,  $T_j$  is the total waiting time of the j-th taxi.

**Step2:** Analysis of decision variables and environmental parameters. Assuming that the taxi is traveling in the same direction and overtaking is not allowed, the straight-line distance from the exit of the plane to the "boarding point" is s, and the total length of the overpass is L. Because the distance between the boarding points is small, it can be ignored. Ensure that the exit of the station, the "boarding point" on the same side, the flyover, and the "boarding point" on the opposite side are on the same straight line and perpendicular to the two parallel lanes. Considering the queue time of passengers and taxis, then:

The waiting time for passengers on the same side is:

$$t_a = \frac{s}{v_{person}} \quad (2)$$

The waiting time for passengers on the opposite side is:

$$t_b = \frac{s+L}{v_{person}} \quad (3)$$

The waiting time for the taxi on the same side is:

$$T_a = \frac{L_1}{v_{vehicle}} \quad (4)$$

The waiting time for the taxi on the opposite side is:

$$T_b = \frac{L_2}{v_{vehicle}} \quad (5)$$

**Step3:** The total time from entering the station to leaving the taxi is:

$$T = T_j + t_i \quad (6)$$

**Step4:** Analysis of constraint conditions. After the gates are opened, if the speed of the taxi is too fast, there will be a certain safety hazard. Therefore, the speed of the taxi should not exceed the maximum speed limit  $V_{max}$  stipulated by the airport.

$$v_{vehicle} \leq V_{max} \quad (7)$$

In summary, the constraints given:

$$\left\{ \begin{array}{l} t_a = \frac{s}{v_{person}} \\ t_b = \frac{s+L}{v_{person}} \\ T_a = \frac{L_1}{v_{vehicle}} \\ T_b = \frac{L_2}{v_{vehicle}} \\ T = T_j + t_i \\ v_{vehicle} \leq V_{max} \end{array} \right. \quad (8)$$

### 2. Solving the model

According to the above analysis, the model of question three is:

$$\left\{ \begin{array}{l} \max \eta = \frac{1}{\sum_{i=1}^a t_i/a + \sum_{j=1}^b T_j/b} \times \frac{2k}{\alpha^k} \\ t_a = \frac{s}{v_{person}} \\ t_b = \frac{s+L}{v_{person}} \\ T_a = \frac{L_1}{v_{vehicle}} \\ T_b = \frac{L_2}{v_{vehicle}} \\ T = T_j + t_i \\ v_{vehicle} \leq V_{max} \end{array} \right. \quad (9)$$

Based on the above analysis, the link formula between the efficiency  $\eta$  and the dependent variable can be obtained. By calculating 0.9 times the maximum number of cars available for searching, the number of ride points at this time can be obtained. Under the conditions of ensuring the safety of vehicles and passengers, in order to maximize the overall riding efficiency, a total of 10 parking spaces are set up, that is, five "boarding points" are set on both sides, and the efficiency of the riding system is at the maximum at this time.

### 2.3 SHORT-DISTANCE TAXI ROUND-TRIP PRIORITY ARRANGEMENT MODEL

In order to balance the revenue of taxis as much as possible, we are now considering giving some priority to taxis returning short-distance passengers and resetting the queuing plan. For this reason, we separately studied the round-trip situation of most taxis between the airport and the city[4], and the round-trip situation of short-distance taxis between the airport and the city, and compared the benefits of the two in the same observation period. With the empty-load rate in the same observation space, search for short-distance taxi positions that make the difference between the revenue and the empty-load rate of the long-distance taxi relatively small as the final queuing plan.

#### 1. Arrange the establishment of the model

The revenue of taxi passengers at the airport is directly related to the mileage (or time) of passengers. In order to balance the revenue of drivers, the following "optimization" plan is determined:

Assuming that a taxi queue waiting time  $T_0$  in a certain period of time, the driving time of passengers returning to the urban area is  $T_1$ , and the income is  $P_1$ ; while the short-distance passenger driving time is  $T_2 < T_1$  and the income is  $P_2 < P_1$ , it needs to return to the airport via  $2T_2$ .

The average return per unit time of a car carrying passengers back to the urban area:  $P_1/(T_0 + T_1)$ , The average return per unit time of a car carrying passengers back to the urban area:  $P_1/(T_0 + T_1)$ . Average unit time revenue of a short-distance passenger-carrying vehicle:  $P_2/(T_0 + T_2)$ .

If the revenue of all taxis is required to be as balanced as possible, that is, the waiting time  $t$  for taxis returning to the airport for a short distance must meet:

$$\min_{t \geq 0} \left[ \frac{P_2 + P_1}{T_0 + 2T_2 + T_1 + t} - \frac{P_1}{T_0 + T_1} \right] \quad (10)$$

The  $T_0, T_1, T_2, P_1, P_2$  can be determined according to

the actual situation of the airport, and the  $t$  value can be obtained. For a short-distance taxi carrying passengers, as long as it returns within 2% of the passenger's departure, and waits for more time, it will be "prioritized" to carry passengers.

Take the airport in the previous question as an example: the average driving from the airport to the city is  $T1=45$ ,  $P1=140$ ; the average short-distance driving is  $T2=10$ ,  $P2=30$ , assuming the waiting time in line  $T0=60$ .

From the above formula,  $t=2.5$  can be obtained, so it is concluded that if the short-distance passenger taxi at the airport can return to the airport within 20 minutes, the car only needs to wait for another 2.5 minutes to "prioritize" departure and carry passengers, and achieve revenue Balanced purpose.

#### 2. "Priority" arrangements

Taxi drivers need to choose between two situations based on personal experience:

(A) Go to the arrival area and wait in line to carry passengers back to the city. Taxis must wait in line at the designated "car storage pool" and enter the venue according to the "first come, first come" line to pick up passengers. The waiting time depends on the number of taxis and passengers in the queue, and a certain time cost is required.

(B) Directly vacate and return to the urban area to solicit customers. Taxi drivers will pay no-load charges and may lose potential passenger revenue.

After confirming that the taxi driver returns for a short distance carrying passengers, a certain "priority" will be given to allow him to arrive at the "priority waiting area" for a short wait, so as to enter the first lane for priority transportation; if the taxi is not When the short-distance passenger time requirement is reached, they will not get "priority" after returning to the airport, but will return to the storage tank and wait in line for pick-up. If both of them enter the same lane at the same time, the taxi that returns short-distance passengers will be given priority, and the vehicle that departs from the storage pool needs to stop and wait, as shown in the following figure:

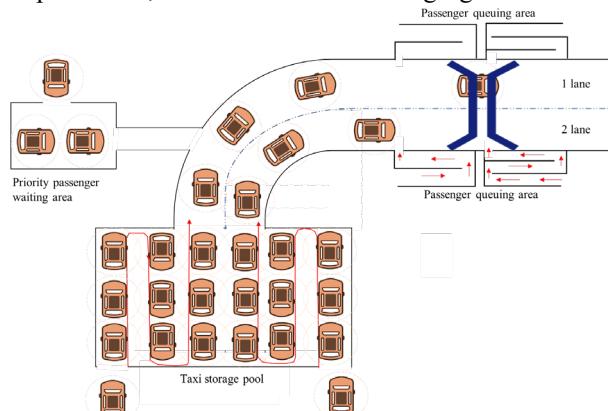


Figure 2 Schematic diagram of airport taxi passenger arrangement 2

### 3. RESULTS AND DISCUSSION

#### 3.1 ADVANTAGES

Through specific and reasonable determination of the airport taxi pick-up mechanism, the ride efficiency under

the number of pick-up points is calculated, and the impact of various environmental parameters on the number of pick-up points is analyzed and discussed, and the airport management department The overall arrangement is suggestive.

### **3.2DISADVANTAGES**

Ignore the influence of seasonal factors on the number of passengers at the airport, only qualitatively analyze the factors that affect decision-making, there is still room for improvement.

Some parameters of the model are estimated based on actual conditions, and there are slight errors in the results.

### **3.3DICUSSION**

This model is based on the planning and processing of taxi pick-up problems in the airport, and can be applied to other types of large transportation hubs at the same time, because they belong to the same type of problem, and the difference is only as long as the flow of people and vehicles and the parameters of some hubs are different, such as large trains. Stations and river, sea ports, etc., can be applied to this model to solve the problem of passenger transfer at the station. In practice, there are many deterministic and uncertain factors that affect decision-making, making it difficult for taxi drivers to make correct judgments. Therefore, airport taxi management departments are required to carry out a reasonable "quantity" release of taxis into the "riding area" based on the actual situation. "As well as arranging a certain number of passengers, providing taxi drivers with a reasonable selection strategy, and improving and balancing the benefits of taxi drivers is of substantial significance.

### **4. CONCLUSIONS**

At the taxi pick-up point of the airport, passengers often queue up when taking a taxi back to the urban area. There are three main types of existing taxi "pick-up points", and the problem is analyzed by two parallel lanes. In order to improve the convenience and overall efficiency of passengers in the process of riding, the "boarding points" are arranged on both sides of the road in turn, and the riding efficiency of the "riding area" is increased to reduce the time wasted due to the front and rear obstruction of taxis. Secondly, based on the safety of vehicles and passengers, to avoid vehicle detention caused by crossing the road, pedestrian overpasses can be built above the lanes on both sides so that the overpass is directly at the exit of the airport, improving the efficiency of passengers

on the opposite side. Considering the overall riding situation, if there are more "boarding points", the vehicles with passengers in the rear will be affected by the vehicle in front, and the riding efficiency will be reduced. Therefore, multiple "boarding points" will be set under the influence of various dependent variables. Pick up point "k, finally, find the optimal arrangement mode of pick up point through the establishment of efficiency model. This question needs to give a certain "priority" to short-distance taxis to balance the short-distance and long-distance taxi drivers' income as much as possible. The passenger carrying income of airport taxis is mainly related to the mileage and time traveled when carrying passengers. In order to make the income among taxi drivers as balanced as possible, the original waiting area queuing method needs to be optimized. According to the actual situation of the airport and the regularity of taxis between the airport and the city, the driving conditions of short-distance and long-distance taxis can be determined, and the income and empty-load ratio of the two types of taxis can be calculated separately, and the income of the rental car driver can be balanced by comprehensive consideration. The goal is to give the optimal short-distance bus "priority" queuing plan.

### **REFERENCES**

- [1] Sun Jian, Modeling and Simulation of Landside Passenger Service Resources in Air Hubs Based on Queuing Theory [D], China University of Mining and Technology (Beijing), 2017.
- [2] Jiang Heng, Research on Arrival and Turnover Parking Spaces of Large-scale Hub Taxi, Urban Transportation Planning Academic Committee of China Urban Planning Society, Fuzhou Municipal People's Government Bus Priority and Congestion Mitigation Countermeasures-China Urban Transportation Planning 2012 Annual Conference and Proceedings of the 26th Academic Symposium, 2012:9.
- [3] Wu Yunbiao, Tang Dafang. Research on airport taxi pick-up plan and pick-up point[J]. Journal of Liaoning University of Technology (Natural Science Edition), 2021, 41(04): 274-280.
- [4] Ma Yanli, Zuo Xuewu, Chu Zhengqing. Analysis of the decision-making model for the optimal allocation of airport taxi drivers[J]. Journal of Hebei North University (Natural Science Edition), 2020, 36(11): 12-23.

# Analysis Of Pollutant Particle Distribution and Content Based on Diffusion Model

*Hongwei Gao<sup>1</sup>, Xin Meng<sup>2</sup>, Xiyan Meng<sup>3</sup>*

<sup>1</sup>*School of metallurgy and energy, North China University of Science and Technology, Tangshan 063210, Association of mathematical modeling, China;*

<sup>2</sup>*School of Basic Medicine, North China University of Science and Technology, Tangshan 063210, Association of mathematical modeling, China;*

<sup>3</sup>*School of mining engineering, North China University of Science and Technology, Tangshan 063210, Association of mathematical modeling, China*

**Abstract:** A mathematical model was established to analyze the impact of nuclear wastewater on the territorial sea environment of China, the United States and Japan. And assess the impact of radioactive substances in nuclear wastewater on the Marine environment, organisms and humans in the future. In this paper, a diffusion model that obeys normal distribution is established by the application of Fick's law, which can intuitively show the concentration distribution of one-dimensional plane. At the same time, according to the characteristics of pollutant diffusion with Markov property, a Markov chain model is established, which can calculate the time required for the completion of diffusion under the condition of limited data.

**Key words:** Fick's Law; Markov Chain Model; Pollution Concentration; Random Diffusion.

## 1. BACKGROUND

Radiation leaks from reactors at the Fukushima Nuclear power plant following a tsunami triggered by the March 11, 2011, earthquake in Japan. In an effort to cool the reactors and avoid a meltdown, Tepco has flooded them with cooling water, creating a growing pool of radioactive waste. Today, more than 1.2 million tons of nuclear waste is stored in 1, 182 huge steel containers, and 140 tons of nuclear waste needs to be stored every day, but these huge containers can't hold any more. In February 2020, the Japanese government said the most realistic way to deal with nuclear waste was to dilute it and release it into the sea or atmosphere. In order to reduce the radioactive substances in nuclear waste water, a kind of equipment called advanced liquid treatment system can use special materials to reduce the concentration of more than 60 radioactive substances, such as strontium and cesium, to a certain standard value, but the radioactive substance tritium is almost impossible to remove.

While some groups say tritium has no real impact, some nuclear experts say it is not true that tritium is the only radioactive substance in waste water. The waste water also contains the radioactive isotope carbon-14, which has a half-life of 5, 370 years and can enter all living things to damage human DNA. Three radioactive isotopes, carbon-14, cobalt-60 and strontium 90, take longer to degrade and enter the Marine food chain more easily. Pang said that if nuclear water is discharged into the sea, given the characteristics of ocean currents, nuclear water will affect Pacific rim countries, especially fisheries and aquaculture,

for decades to come. Vladimir Rakov, a PhD biologist and researcher at the Russian Academy of Sciences, said that even if nuclear waste water is carefully cleaned and discharged into the ocean, it could still cause radioactive isotopes to be retained in Marine life, including fish, and then accumulate in humans. On April 13, 2021, after a cabinet meeting, the Japanese government officially decided to discharge 1.23 million tons of nuclear waste water from nuclear reactors into the Pacific Ocean within about 10 years. The move drew widespread attention and protest from the international community.

## 2. THE ESTABLISHMENT AND SOLUTION OF MODEL

Fick's law is a law that describes the relationship between diffusion flow and diffusion material concentration gradient in a diffusion system. It is intended to further understand Fick's law through the analysis of the migration and transformation mechanism of pollutants in water, so as to obtain the migration and transformation law of pollutants in water. Based on the theories of stochastic process, Brownian motion and geometric Brownian motion, the continuous model is approximated by random walk, and the diffusion of pollutant concentration conforms to geometric Brownian motion. These two results are the theoretical basis of water quality migration model.

The pollutants entering water can be divided into two categories: conserved pollutants and non-conserved pollutants. Conservation of pollutants refers to the pollutants enter the water, distribution of pollutants, the concentration of the pollutants in the water change with the change of time, but the total amount of pollutants, attenuation, also is not going to happen some polymer such as heavy metals, organic compounds, these pollutants degradation or doesn't happen very hard, these pollutants may spread will not happen, There is little harm to humans temporarily, but in the long run, large amounts of accumulation in the environment without degradation can lead to serious harm.

Non-conserved pollutants refer to the degradation of pollutants in water, and their distribution changes over time. At the same time, their total amount also changes, such as the decomposition of microorganisms in water, the biochemical reaction between pollutants and pollutants, and the oxidative decomposition of oxygen-consuming organic matter.

The migration and transformation process of pollutants in water body takes point source as an example. When point source sewage is discharged into water body, the migration and transformation includes convection, diffusion, biodegradation, suspension and sedimentation. In order to facilitate the study, it is divided into physical process, chemical process and biological process, among which the physical process is the most important, because in practical problems, there can be no chemical process and biological process, but the physical process must exist. Physical process is divided into convection and diffusion, diffusion can be divided into molecular diffusion, turbulent diffusion, floating and dispersing action. The principle of turbulent diffusion mainly comes from turbulence. Turbulent flow and laminar flow do not exist in the constant change of velocity, turbulent flow field particle instantaneous value is equivalent to the average random pulsation caused by the dispersion phenomenon. Dispersion is also known as hydrodynamic dispersion or hydraulic dispersion. The thermal kinetic energy of the particle and the convection of the fluid are the result of molecular diffusion and mechanical mixing. So diffusion has both molecular diffusion and mechanical diffusion. According to the kinetic theory of molecules, the number of molecular collisions per unit time is huge. Under normal conditions, the number of collisions per liter of volume per second is as high as more than 1032 times, indicating that the molecules are constantly making irregular movement. The interpenetration of two different substances through molecular motion is called molecular diffusion. Molecular diffusion satisfies Fick's law. The flux of diffusion is proportional to the gradient of concentration in this direction.

We consider only the molecular diffusion of conserved pollutants. Through existing in the static water contamination particle radial diffusion, and diffusion velocity in different directions of the same ideal conditions is calculated in the diffusion model of the spread of nuclear waste water, the concentration of radioactive particles, and other performance it spread of the data, to the nuclear pollution impact on environment of the territorial sea, a concentration of pollutant particles directly decide the premise. According to Fick's law (describing the relationship between diffusion flow and diffusion material concentration gradient in diffusion system), an instantaneous point source one-dimensional diffusion diagram model is established. The radioactive pollutant particles in sewage are regarded as a whole [2] and taken as the object of study to analyze the characteristics of their movement (diffusion), which conforms to the molecular diffusion process caused by Brownian motion. It is a random molecular diffusion process affected by concentration gradient and finally tends to be stable. According to the characteristics of the diffusion process, all radioactive particles as a collection as a whole, the diffusion process as a movement tends to smooth the process, the spread of the particles over time because of pollution situation with markov [3] sex, namely the spread of the state of the next moment only by the diffusion of the moment on a relevant, has nothing to

do with the history before a moment. Markov model was established by taking the stable state achieved by diffusion completion as absorption state and the diffusion direction of pollutant particles as probability event  $\{X(n), n = 0, 1, 2 \dots\}$ . The space of is E, If any  $m(m \geq 0)$  integers  $n_1, n_2, n_3 \dots nm(0 \leq n_1 \leq n_2 \leq n_3 \leq \dots \leq nm)$  and any natural numbers k and  $i_1, i_2, i_3 \dots im, j \in E$  are satisfied  $P\{X(nm + k) = j | X(n_1) = i_1, X(n_2) = i_2, \dots, X(nm) = im\} = P\{X(nm + k) = j | X(nm) = im\}$ , then  $X(n)$  is Markov chain. It is assumed that the water body is stationary and the pollutants start from point source diffusion. One-dimensional diffusion diagram of instantaneous point source is simulated as follows: Pollutant concentration is assumed to be  $C(x, t)$ , the length direction of concentration distribution is x direction, and the pollution source is put in  $x=\sigma$ . Fick's law is used as a complement to obtain the following results:

Hypothesis stationary, water pollutants begins with point source diffusion, the instantaneous point source spread one dimensional figure simulation is as follows: a pollutant concentration to  $C(xt)$ , the length of the concentration distribution is in the x direction, pollution sources on the  $x = \sigma$ , based on Fick's law, the concentration of do the Fourier transform expression, it is concluded that the diffusion concentration and diffusion distance of normal distribution.

Assume the one-dimensional diffusion diagram model of instantaneous point source as  $C(x, t)$ , the length direction of the concentration distribution is x, and the pollution source is placed at  $x=\sigma$ . Then, the Normal distribution relationship between diffusion concentration and diffusion distance is obtained by Fourier transform of the concentration expression.

$$\bar{c}(a, t) = \frac{M}{\sqrt{2\pi D t}} e^{-\frac{(x-\sigma)^2}{4Dt}}$$

As the diffusion of polluting particles over time is markov [5], that is, the diffusion state at the next moment is only related to the diffusion situation at the last moment, and has nothing to do with the historical situation before the last moment. Markov model was established by taking the stable state achieved by diffusion completion as absorption state and the diffusion direction of pollutant particles as probability event

The Markov model is established and the normal distribution function of the concentration range relationship at different moments in the first question is integrated and averaged to obtain the one-dimensional average concentration at different moments [2], as shown in Table 1:

Table 1 Average concentration

$t_1$	$t_2$	$t_3$	$t_4$
$c_1$	$c_2$	$c_3$	$c_4$

Each radioactive particle is studied and its complex molecular motion is divided into three categories:

Scenario A: Movement away from point sources of contamination

Case B: Not moving

Case C: Movement near the point source of contamination

According to the average concentration at different times, the function relation can be established as follows:

$$P_{A_n} = \frac{\frac{1}{C_{n+1}} - \frac{1}{C_n}}{\frac{1}{C_{n+1}}}$$

$$P_{B_n} = \frac{\frac{1}{C_n} - \frac{1}{C_{n+1}}}{\frac{1}{C_{n+1}}}$$

$$P_{C_n} = 0$$

The transfer matrix is established as follows:

$$\begin{matrix} R & A & B & C \\ A & \left( \begin{array}{ccc} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{array} \right) \\ B & \left( \begin{array}{ccc} P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{array} \right) \\ C & \left( \begin{array}{ccc} P_{31} & P_{32} & P_{33} \end{array} \right) \end{matrix}$$

$$\left\{ \begin{array}{l} P_{11} = \frac{P_{A2}}{P_{A1}} \\ P_{22} = \frac{P_{B2}}{P_{B1}} \\ P_{12} = 1 - P_{11} - P_{13} \\ P_{21} = 1 - P_{22} - P_{23} \\ P_{13} = P_{23} = P_{33} = P_{32} = P_{31} = 0 \end{array} \right.$$

Subtract R from the identity matrix and inverse:

$$(I - R)^{-1}$$

$$\left( \begin{array}{ccc} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{array} \right)$$

$$\left\{ \begin{array}{l} t_{12} = 1 - P_{11} - P_{13} \\ t_{21} = 1 - P_{22} - P_{23} \\ t_{13} = t_{23} = t_{33} = t_{32} = t_{31} = 0 \end{array} \right.$$

$t_{12} + t_{23}$  denoting the time to complete diffusion is denoted as  $t_m$

The diffusion range can be obtained by substituting  $t_m$  into the positive distribution graph/function.

### 3. MODEL TO PROMOTE

Aiming at the uncertain mathematical model, it mainly applies the stochastic method and uses the probabilistic distributed random variable to express the uncertainty. From this point of view, the model used in this topic is suitable for the occurrence of things with Markov property. For models with recurrent states in state space, such as the

motion model of a finite number of molecules [6], the locomotive model with a finite number of starting and ending points, and the human health and disease transfer model, the Markov chain model used in this question can be applied in these cases.

### 4. EVALUATION AND IMPROVEMENT OF THE MODEL

#### 4.1 EVALUATION OF THE MODEL

##### • Advantages

(1) In the case of less data and known information, the model still has the characteristics of relatively accurate grasp of the space transfer of radioactive particles, and calculates the relatively accurate transfer time of radioactive pollutants.

(2) The global territorial sea is regarded as a decentralized system, and the diffusion of each radioactive particle is tracked. The overall use of macro ideas, but the micro calculation is also very precise accuracy, making this model universal and scientific.

(3) A large number of mathematical and physical formulas are used, and variable symbols are also explained in detail. The whole reasoning process is easy to understand, the logic is clear, and the data is concise and intuitive.

(4) Compared with grey prediction, the model applied in this case has higher prediction precision and more accurate results.

(5) This method has good effect on process state prediction and can be used to predict the dangerous state of production site.

Be able to calculate the probability of a system with maintenance capability and multiple degradation states

##### • Disadvantage

(1) the influence of ocean current movement and tidal movement of sea water on the diffusion of radioactive pollutants is not considered.

(2) Taking only the region as the research object, pollution is a dynamic and continuous process, and it is not enough to directly reflect the accurate situation of pollution in territorial sea based on concentration changes in different time periods.

(3) Is not applicable to system medium and long term forecast. It is assumed that the probability of state change is fixed, whether for failure or maintenance

(4) All events are statistically independent, so the future state is independent of all past states, unless the two states are closely aligned

(5) You need to understand the various probabilities of state change

(6) The knowledge of matrix operation is complicated and difficult for non-specialists to understand

#### 4.2 Model improvement

From geographical knowledge, Japan's territorial sea is the intersection of two ocean currents - Kuril Island cold current and Japan Warm Current, but also the start of the Pacific warm current, the main currents have equatorial circulation, subtropical circulation, sub-polar circulation. In this way, the velocity of the ocean current can be used as the velocity of the ocean body of water; The direction of ocean current is taken as the main direction of particle motion. Based on this, the above two points can

supplement the original model.

REFERENCE

- [1] Guo A. Numerical simulation of the influence of drool dynamics and sewage diffusion in the Yangtze River [D]. Hohai University, 2007.
- [2] Wu Jin, Wu Shengxing. Random Model of Chloride ion concentration on reinforced Bar surface of Concrete in Marine Environment [J]. Journal of Hohai University (Natural Science Edition), 2004(01):38-41.
- [3] Yang Mei-mei. Study on the Diffusion transport model of oil-bearing wastewater and its application in the Development process of South China Sea oilfield [D]. Ocean University of China, 2007.
- [4] Jiang Fuhou. Study on pollutant migration based on stochastic diffusion process [D]. Huazhong University of Science and Technology, 2012.
- [5] Shao Xuejun, Xia Zhenhuan. Journal of Hydraulic Engineering, 1990(10):32-40.
- [6] Ren Tianshan, CHENG Jianping, ZHU Li, et al. Environment and Radiation [M]. Beijing: Atomic Energy Press, 2007.

# The Catalytic Process of C4 Olefin Preparation from Ethanol was Investigated based on Particle Swarm Optimization

Zixuan Ge<sup>1\*</sup>, Zhenbin Zhang<sup>2</sup>, Bingbing Shi<sup>3</sup>

<sup>1</sup> School of Electrical Engineering, North China University of Technology, Tangshan 063210, China;

<sup>2</sup> College of Science, North China University of Science and Technology, Tangshan 063210, China;

<sup>3</sup> School of Artificial Intelligence, North China University of Technology, Tangshan 063210, China;

\*Corresponding Author.

**Abstract:** Ethanol is a key renewable compound. It is of great value to study the relationship between ethanol conversion and C4 olefin selectivity and temperature, and to design a reasonable combination of catalysts to improve the yield of C4 olefin. In this paper, the effects of different catalyst combinations and reaction temperature on the yield of C4 olefin were studied, and the optimal reaction conditions were sought to improve the yield of C4 olefin. Finally, a complete experiment was designed. We established a particle swarm optimization algorithm model based on linear transformation, and obtained the conditional optimal value of catalyst combination and temperature with the highest C4 olefins yield as possible under different constraints. The optimal value of catalyst performance was obtained. In order to ensure the integrity and reliability of the experiment, the optimal solution catalytic combination of the third question was used to conduct a more detailed study at the temperature between the regions and design experiments to study the temperature at which the catalytic efficiency was higher. We designed three groups of comparison experiments and two groups of optimization experiments according to the actual conditions.

**Keywords:** Olefin Preparation From Ethanol; Particle Swarm Optimization Algorithm; Comparative Optimization Experiment

## 1. PROBLEM ANALYSIS

According to the linear relationship model of different catalyst combinations and temperature on ethanol conversion and C4 olefin selectivity, a particle swarm optimization algorithm model based on linear transformation was established, and the conditional optimal value of catalyst combinations and temperature with the highest C4 olefin yield under different constraints was obtained.

To design five groups of new experiments, it is necessary to consider that the new experiment and the original experiment are evenly distributed and have good comparison and optimization. First, the variable data of each catalyst combination was examined, and the combinations lacking contrast were examined. A new experiment was added to determine whether the charging method was related to the catalyst combination. Then select the two factors with the maximum weight solved in the second question, control the single variable, add two

experiments of the second charging method, form the comparison experiment within the same charging method, and increase the accuracy of the analysis results of the second charging method; In addition, two optimized temperature experiments were added, and no calculation and analysis was performed at low temperature with low ethanol conversion rate and low C4 olefins selectivity, and at high temperature with missing data. In this way, the catalytic combination of the optimal solution of the third question can be used to study how much temperature the catalytic efficiency is higher and more accurate in a certain range of temperatures.

## 2. MODEL ASSUMPTIONS

In order to establish a simplified model, the team simplified each participant of the model as follows during the modeling process:

1. In the process of preparing C4 olefin by ethanol coupling, all the experimental processes are legal and standardized, and no safety accidents will occur.
2. It is assumed that the density of materials does not change during the preparation of C4 olefin by ethanol coupling.
3. It is assumed that there are no other side reactions during the preparation of C4 olefin by ethanol coupling, and the utilization rate of carbon atom is 100%.
4. It is assumed that the reaction vessel used in the preparation of C4 olefins by ethanol coupling is of the same size and quality.
5. The quality changes before and after the reaction of preparing C4 olefin by ethanol coupling are not considered.
6. Assume that the design environment of this paper is a real laboratory environment.

## 3. DEFINITION AND NOTATION

In order to simplify problem analysis and data processing, symbols are stipulated as follows:

symbol	define
$\alpha$	Ethanol conversion
$\beta$	C4 olefins selectivity
$T$	Reaction temperature
$t$	The reaction time
$s_i$	Catalyst combination variable
$\theta$	C4 olefin yield
$n$	The number of particles
$c_1$	Individual learning factors of particles, also known as individual accelerators
$c_2$	The particle's social learning factor, also known as the

$\omega$	social acceleration factor
$v_i^d$	Inertial weight of velocity
$x_i^d$	The velocity of the ith particle in the DTH iteration
$f(x)$	The position of the ith particle in the DTH iteration
$pbest_i^d$	Fitness value at position X (generally taking the value of the objective function)
$pbest_i^d$	By iteration d, the ith particle goes through the best position
$gbest^d$	By iteration d, the best position that all particles have passed through

## 4.ESTABLISHMENT AND SOLUTION OF PSO MODEL

### 4.1 PREPARATION OF PSO MODELS

We choose the PSO model of multi-objective optimization, namely particle swarm optimization algorithm. We not only change the range of inertia coefficient in the velocity of the iterative particle, so as to avoid too fast convergence. At the same time, the linear model and PSO model are combined effectively, and a particle swarm optimization algorithm based on linear transformation is proposed, which makes the inertia coefficient decrease with the increase of the number of iterations, so as to get the global optimal situation as far as possible<sup>[1]</sup>. Finally, by obtaining the objective function (C4 olefin yield as high as possible) in line with the constraint conditions, the catalyst combination and temperature under the condition of C4 olefin yield as high as possible were obtained.

Particle Swarm Optimization (PSO) is a Swarm cooperative search algorithm developed by simulating the foraging behavior of birds. In this section, it is used to solve the selected catalyst combination and temperature with the highest C4 olefin yield possible under the same experimental conditions.

In particle swarm optimization, the velocity of the particle at each iteration  $v_i$  And the location  $x_i$  The update formula is as follows:

$$v_i^d = wv_i^{d-1} + c_1 r_1 (pbest_i^d - x_i^d) + c_2 r_2 (gbest^d - x_i^d) \quad (1)$$

$$x_i^{k+1} = x_i^k + av_i^{k+1} \quad (2)$$

Among them,  $\omega$  is inertial weight,  $c_1$  and  $c_2$  is the learning factor,  $r_1$  and  $r_2$  Is a random number between [0, 1].  $r_1, r_2$  in (0, 1) Between randomly.  $pbest$  represents the individual extremum, which records the best location where a single particle is found.  $gbest$  Represents the total extremum that records the best position experienced by the entire population.

### 4.2 LINEAR TRANSFORMATION MODEL PRINCIPLE

Suppose the C4 olefins yield is  $\theta$ . Under the constraint conditions, the ultimate goal is to seek the highest C4 olefin yield. The objective function is as follows:

$$\max \theta = \alpha \sqcup \beta \quad (3)$$

$\alpha$  represents ethanol conversion and  $\beta$  represents C4 olefin selectivity.

To solve this objective function, the relationship between catalyst combination and temperature and ethanol conversion and C4 olefins selectivity was studied. Reference, specific constraints are as follows:

In the case of charging mode I regardless of temperature:

$$\begin{aligned} s.t. & \left\{ \begin{array}{l} 0.1 \leq s_1 \leq 200 \\ 0.1 \leq s_2 \leq 5 \\ 0.1 \leq s_3 \leq 200 \\ 0.1 \leq s_4 \leq 1.68 \\ 250 \leq T \leq 400 \end{array} \right. \\ & \alpha = 4.8628 - 0.0867s_1 - 0.0010s_2 + 0.2485s_3 - 0.1810s_4 + 0.5545T \\ & \beta = -0.9429 + 0.2566s_1 - 0.1714s_2 - 0.1250s_3 + 0.0732s_4 + 0.2933T \end{aligned} \quad (4)$$

Regardless of temperature, charging mode II:

$$\begin{aligned} s.t. & \left\{ \begin{array}{l} 0.1 \leq s_1 \leq 200 \\ 0.1 \leq s_2 \leq 5 \\ 0.1 \leq s_3 \leq 200 \\ 0.1 \leq s_4 \leq 1.68 \\ 250 \leq T \leq 400 \end{array} \right. \\ & \alpha = 357.6536 + 0.1238s_1 - 7.3310s_2 - 0.0389s_4 + 0.4272T \\ & \beta = 0.0590s_1 + 0.00005 - 0.0212s_4 + 0.2730T \end{aligned} \quad (5)$$

In the case of temperature below 350 degrees and charging mode I:

$$\begin{aligned} s.t. & \left\{ \begin{array}{l} 0 \leq s_1 \leq 200 \\ 0 \leq s_2 \leq 5 \\ 0 \leq s_3 \leq 200 \\ 0 \leq s_4 \leq 1.68 \\ 250 \leq T \leq 350 \end{array} \right. \\ & \alpha = 4.8628 - 0.0867s_1 - 0.0010s_2 + 0.2485s_3 - 0.1810s_4 + 0.5545T \\ & \beta = -0.9429 + 0.2566s_1 - 0.1714s_2 - 0.1250s_3 + 0.0732s_4 + 0.2933T \end{aligned} \quad (6)$$

In the case of charging mode II at temperatures below 350 degrees:

$$\begin{aligned} s.t. & \left\{ \begin{array}{l} 0 \leq s_1 \leq 200 \\ 0 \leq s_2 \leq 5 \\ 0 \leq s_3 \leq 200 \\ 0 \leq s_4 \leq 1.68 \\ 250 \leq T \leq 350 \end{array} \right. \\ & \alpha = 357.6536 + 0.1238s_1 - 7.3310s_2 - 0.0389s_4 + 0.4272T \\ & \beta = 0.0590s_1 + 0.00005 - 0.0212s_4 + 0.2730T \end{aligned} \quad (7)$$

### 4.3 ESTABLISHMENT OF JOINT MODEL

The results show that when the inertia coefficient  $\omega$  is large, the global convergence capability is strong, and when the inertia coefficient  $\omega$  is small, the local convergence capability is strong. Therefore, in the process of iteration, if the inertia coefficient  $\omega$  can be linearly changed with the number of iterations, the initial inertia coefficient  $\omega$  value will be relatively large and the global search ability will be strong. And with the increase of the number of iterations, the inertia coefficient  $\omega$  slowly decreases, so as to ensure that particle swarm optimization algorithm has relatively strong local search ability when entering the optimal region. The formula for calculating the inertia coefficient  $\omega$  linear transformation is as follows:

$$\omega = (\omega_1 - \omega_2) \frac{MaxDT - DT}{MaxDT} + \omega_2 \quad (8)$$

Among them,  $\omega_1/\omega_2$  Is the initial and final values of the inertia coefficient,  $MaxDT/DT$  Is the maximum number of iterations and the current number of iterations.

According to PSO model and linear transformation model, the particle swarm change based on linear transformation is established<sup>[2]</sup>.

Based on the above new calculation formula of inertia coefficient and PSO model, the algorithm flow obtained by us is updated as follows:

Step 1: Within the constraint conditions, the particle swarm is initialized and particles are generated randomly  $pbest$  \  $gbest$ .

Step 2: Calculate the particle position, velocity and inertia coefficient after each iteration according to the formula.

Step 3: Calculate the fitness from the fitness function and  $pbest$  Compare and select the best result.

Step 4: Judge  $pbest$  and  $gbest$  results, select the optimal.

Step 5: Judge whether the termination conditions are met, if so, output the results; If not, go back to step

#### 4.4 JOINT PARTICLE SWARM OPTIMIZATION MODEL

After calculation, without considering temperature:

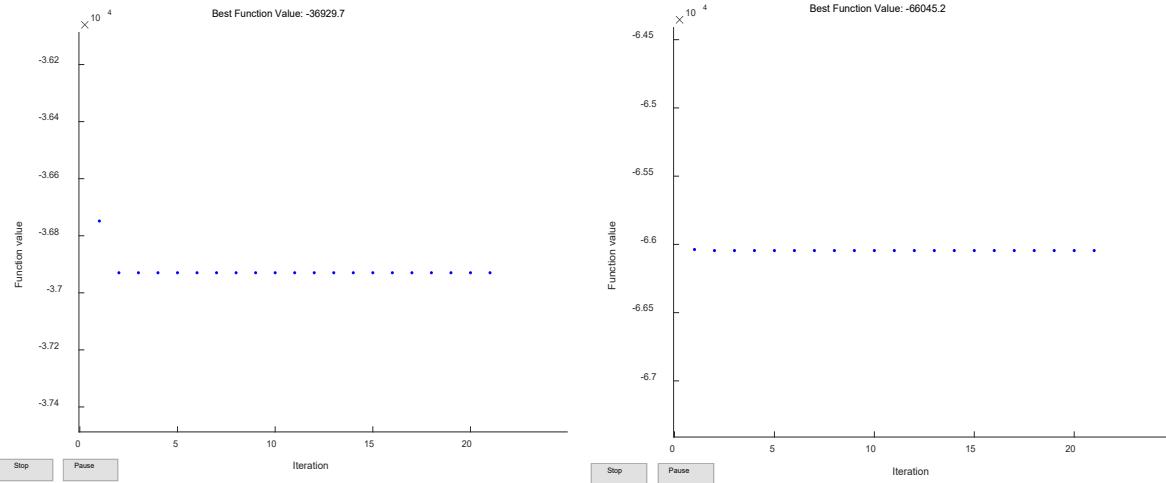


Figure 1. Variation of the optimal function value with the number of iterations

Under the loading mode I (the first left of Figure. 1), when  $s_1 = 200, s_2 = 0.1, s_3 = 200, s_4 = 1.68, T = 400$ , the yield of C4 olefins reaches the maximum. Under charging mode II (first from the right of Figure. 1), when  $s_1 =$

$200, s_2 = 0.1, s_3 = 197, s_4 = 1.68, T = 400$ , the yield of C4 olefins reaches the maximum.

In the case of temperatures below 350 degrees:

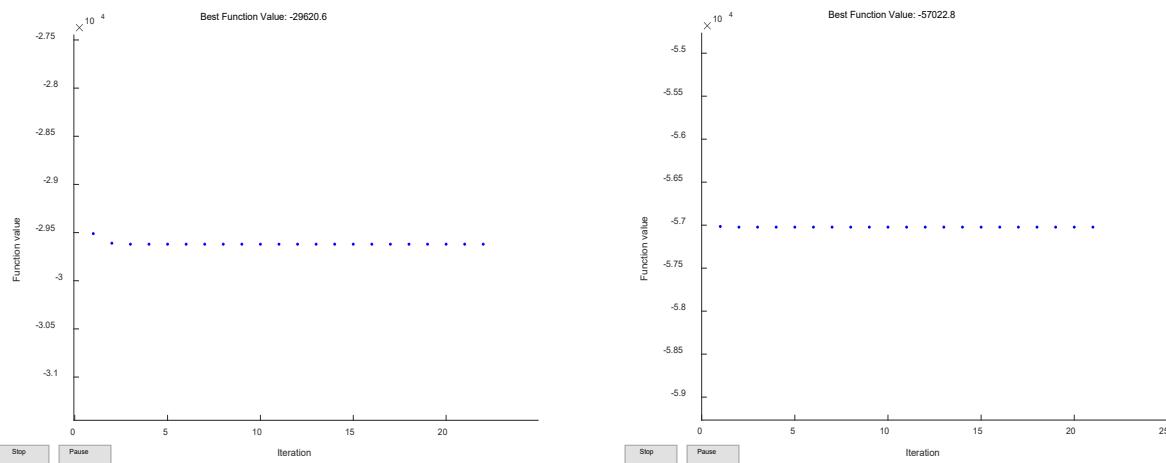


Figure 2. Variation of the optimal function value with the number of iterations

Under the loading mode I (the first left of Figure.2), when  $s_1 = 200, s_2 = 0.1, s_3 = 200, s_4 = 1.68, T = 400$ , the yield of C4 olefin has the maximum value. The maximum C4 olefin yield is obtained when  $s_1 = 200, s_2 = 0.1, s_3 = 198, s_4 = 1.68, T = 400$  under charging mode II (first from the right of Figure.2). Under these conditions, Co is highly dispersed on the catalyst surface, and the catalyst surface is suitable for acid and base, so it has a soft high catalytic activity. The conversion rate of ethanol is 84%, and the selectivity of C4 ene is 53%.

In conclusion, when the mixing ratio of Co/SiO<sub>2</sub> to HAP is 1:1, the reaction temperature is 400°C, and the Co load is 1 wt%, the catalyst performance is optimal.

#### 5.SUPPLEMENT OF EXPERIMENT

##### 5.1 CONTRAST EXPERIMENT

###### 5.1.1 COMPARATIVE TESTS UNDER DIFFERENT

#### CHARGING METHODS

According to data analysis, the catalyst combination mode of A9 and B5 and A12 and B1 is the same, but there are few differences in experimental data due to different charging methods. In order to determine whether the charging method affects the performance of the catalyst, we designed a set of experiments based on this. According to the catalyst combination of the optimal solution of the third question, the charging method was changed to B charging method, controlling a single variable, while other components remained unchanged<sup>[3]</sup>.

The results were observed at different temperatures, and if the results were found to be approximately the same as the loading method of A, it was believed that the loading method did not affect the effect produced by the catalyst. Otherwise, it is considered that the charging mode affects

the effect of the catalyst, and the temperature is combined with the optimal solution.

Table 1 Comparison between new experimental combination B8 and original combination A4

Serial number	<i>s</i> <sub>1</sub>	<i>s</i> <sub>2</sub>	<i>s</i> <sub>3</sub>	<i>s</i> <sub>4</sub>	T
A4	0.5	200	200	1.68	250
					275
B8	0.5	200	200	1.68	300
					325
					350

### 5.1.2 THE SAME GROUP COMPARISON TEST WITH B CHARGING MODE

Further study of the data shows that different factors of catalysts have an impact on the conversion rate of ethanol and the selectivity of C4 olefin. According to the weights analyzed in question 2, the weight factors that have a greater impact on the conversion rate of ethanol and the selectivity of C4 olefin are respectively selected in this experiment: The Co loading capacity and the mass of Co/SiO<sub>2</sub> were compared in the same group, and the catalyst combination with higher C4 olefin yield was selected as the original combination for comparative analysis.

Table 2 Comparison of different Co loads between new experimental combination B9 and original combination B1

Serial number	<i>s</i> <sub>1</sub>	<i>s</i> <sub>2</sub>	<i>s</i> <sub>3</sub>	<i>s</i> <sub>4</sub>	T
B1	1	50	50	1.68	250
					275
B9	0.5	50	50	1.68	300
					325
					350

Table 3 Comparison of Co/SiO<sub>2</sub> quality between the new experimental combination B10 and the original combination B6

Serial number	<i>s</i> <sub>1</sub>	<i>s</i> <sub>2</sub>	<i>s</i> <sub>3</sub>	<i>s</i> <sub>4</sub>	T
B6	1	75	75	1.68	250
					275
B10	1	200	200	1.68	300
					325
					350

### 5.2 OPTIMIZATION EXPERIMENT

The experimental temperature range was optimized. According to the data analysis, when the temperature was lower than 350°C, the ethanol conversion and C4 olefins selectivity were lower in all catalyst combinations. When the temperature is higher than 450°C, the data is not sufficient; When the temperature is between 350°C and 450°C, the conversion rate of ethanol and the selectivity of C4 olefin increase rapidly, and the change of the conversion rate of ethanol and the selectivity of C4 olefin cannot be accurately judged. Therefore, the experimental temperature interval of the new experiment is 20°C, and the range is between 350°C and 450°C. The catalyst combination of the new experiment is provided according to the combination type of the optimal solution of the third

question.

Table 4 Comparison of two new experiments at 350°C-450°C

Serial number	<i>s</i> <sub>1</sub>	<i>s</i> <sub>2</sub>	<i>s</i> <sub>3</sub>	<i>s</i> <sub>4</sub>	T
B11 optimal solution	1	75	75	1.68	350
					370
B12 optimal solution	1	200	200	1.68	410
					430
					450

### 6.EVALUATION AND IMPROVEMENT OF THE MODEL

**Advantages:** PSO model is easy to solve continuity problems and can independently and effectively optimize system parameters, so as to approach the optimal solution more quickly, which is suitable for real-valued processing.

**Disadvantages:** PSO model has poor local optimization ability, and tends to produce premature convergence when dealing with multi-peak problems. It has strong robustness and can optimize the model well, while the combination with insufficient sample size or even no sample can not be inferred well<sup>[4]</sup>.

**Improvement:** With the expansion of application scope, PSO algorithm has some problems to be solved, such as premature convergence, dimension disaster, and easy to fall into local extremum.

Adjust the parameters of PSO to balance the global detection and local mining capabilities of the algorithm. Different types of topologies can be designed to change the learning patterns of particles and thus improve the diversity of populations.

PSO is combined with other optimization algorithms (or strategies) to form a hybrid PSO algorithm.

Niche technology can be adopted to further optimize the PSO algorithm.

### REFERENCES

- [1] Liu Yancong, LI Na, LU Jiangyin. Preparation of low carbon olefin from ethanol over P/HZSM-5 Zeolite catalysts [J]. Natural gas chemical industry (C1 chemistry & chemical industry), 2011, 36(02):32-35.
- [2] Qiao Xu, Min Han, Et al., Multi-objective optimization model of converter alloy addition based on particle swarm optimization algorithm, Control and Decision, Vol.25, No.12:1901-1904, 2010.
- [3] Chen Shaoxing, Research on Particle Swarm Optimization and Its Application for Multi-objective Optimization, Dalian: Dalian University of Technology, 2007.
- [4] Si Shoukui, Li Xi, Mathematical Modeling Algorithm and Application, Beijing: National Defense Industry Press, 2011.

# An analysis of the influence of American presidential candidates on American economy

Xueying Jing<sup>1,\*</sup>, Zehui Dou<sup>2,3</sup>, Bingbing Shi<sup>3</sup>

<sup>1</sup>School of Science, North China University of Science and Technology, Tangshan 063210, China;

<sup>2</sup>Engineering Computing and Simulation Innovation Lab, North China University of Science and Technology, Tangshan 063210, China;

<sup>3</sup>College of Artificial Intelligence, North China University of Science and Technology, Tangshan 063210, China;

\*Corresponding Author.

**Abstract:** Considering from the dimensions of presidential politics, party politics, and political cycles, this US election is of key significance. The elected people's views on Sino-US economic and diplomatic relations are worth studying. First, cluster analysis of 7 impact indicators to obtain key factors in five aspects: finance, environment, external affairs, people's livelihood, and medical care; secondly, through GM (1, 1) prediction, we can get Trump's re-election in 2021. The quantity is 21, 644, and then through multiple linear regression, Trump's proposition on the US economy and the total economic value are determined to be 22, 294; using Biden's 8-year vice presidential tenure data to obtain his proposition and the total economic value is 22752.7. It quantitatively shows that Trump's re-election economic growth rate was small, and development was sluggish; after Biden was elected, he updated his policy propositions, the total economic growth rate was relatively large, and the US economy developed healthy.

**Key words:** Multiple Linear Regression Analysis; Grey Prediction Model; Economic Impact; Epidemic Situation

## 1. INTRODUCTION

The 2020 U.S. election has attracted attention from all walks of life in the world because it will affect the domestic and foreign affairs of the United States in the next four years. Judging from the current developments in the general election, Trump has special characteristics different from traditional political figures and the special impact of the new coronavirus pneumonia during the period. As a "non-open" competition between Republican incumbents and Democratic challengers, candidates of the two parties are engaged in financial and trade, economic and financial governance, and other different key development areas (such as COVID-19 combat measures, infrastructure, Taxation, environmental protection, medical insurance, employment, trade, immigration, education, etc.) have different political positions and administrative programs.[1] Under the constraints of current political, ecological, economic and other factors, the influence of either candidate on the economic structure of China, the United States and the world, as well as the predictability of the China policy are the focus of investigation.

## 2. EXPERIMENTAL

### 2.1 INDEX DATA PREPROCESSING BASED ON CLUSTER ANALYSIS

#### 1. STABLISH AN EVALUATION INDEX SYSTEM

After the thrilling "Super Tuesday" and the small "Super Tuesday" that followed, candidates for the 2020 US general election have gradually surfaced. As the presidential candidate launched by the Republican Party, it is a conclusive fact that Trump won in his party, while the Democratic Party introduced Biden as the presidential candidate. The 74-year-old Trump is the 69th US presidential candidate. The candidates of the two parties together are over 150 years old, which means that it has become an inevitable fact that the US economy is in the hands of over-aged leaders.

Looking at all aspects of Trump's administration, the attitude towards the polarization between the rich and the poor and the class gap is very different, and the implementation of the continuous tax cuts will shape the dividend coverage. At the same time, it must be admitted that during his tenure, Trump handed over a relatively beautiful economic answer sheet. From 2017 to 2019, the U.S. economy has shown a continuous growth trend at 2.2%, 2.9% and 2.3%, even if it is affected by 2020. With the impact of the new crown pneumonia epidemic at the beginning of the year, the most pessimistic forecast for the U.S. economy in 2020 has dropped to 1.9%, resulting in a steady growth of 2.3% during the four-year rule. In addition, in terms of jobs and employment rate, Trump added 7 million jobs in the United States during his administration, and the 3.5% unemployment rate reached a record low in 50 years[2][3].

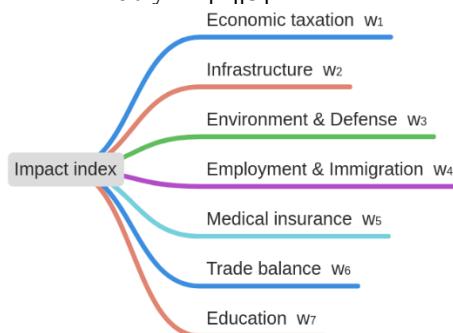


Figure 1 Index evaluation system

Candidates of the two parties differ in financial trade, economic and financial governance, and other different key development areas (such as COVID-19 combat measures, infrastructure, taxation, environmental protection, medical insurance, employment, trade, immigration, education, etc.) Political position and

administrative program. In order to analyze the impression that different candidates might have on the United States, by consulting data and literature, evaluation indicators are selected in economic construction, medical construction, education investment, environmental protection investment, etc. to establish an index evaluation system, as shown in Figure 1.

The above indicators are based on 10 years of data from 2010 to 2019. Among them, the economic tax index mainly refers to the public goods provided by the US government to meet the social needs, as well as the fiscal revenue from the compulsory distribution of free social products. In terms of infrastructure, the U.S. government provides material engineering facilities for social services and public services for residents. This paper takes the government investment in transportation construction as the research object. Biden promised to build a complete clean energy system over the next 30 years. Environmental protection is the focus of government investment in clean energy development. Employment work focuses on the number of jobs and the increase in capital investment employment.

## 2. ABSOLUTE DISTANCE BETWEEN SAMPLE POINTS

The indicator samples have certain similarities. For example, medical investment includes medical insurance, medical setting, medical subsidy, etc., among which the medical setting fund investment is linked to the construction of social infrastructure, which is an indispensable condition for maintaining the stable development of society, and many more. They should be classified using a quantitative method, and the degree of similarity between samples of indicators that affect the economy should be described using quantitative magnification. A thing usually requires multiple variables to describe it. Simplify it as an independent point in the space. The point and the point exist alone, but there is a certain connection, which is not an absolute independent division.

Record the amount invested in a certain indicator as  $v_i (i=1, 2, \dots, 10)$ , and use the absolute distance to measure the similarity between the points in the sample:

$$d(x, y) = \sum_{k=1}^{10} |x_k - y_k| \quad (1)$$

The specific data values of the indicator sample points are shown in Table 1. All the data have the same dimension, and there is no such problem as the large difference in the range of measurement value variation.

Table 1 Sample data (unit: USD 1 billion)

Year	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$
2010	2205	610	324	124	769	150	180
2011	2326	621	347	157	799	178	370
2012	2509	650	360	248	831	190	360
2013	2823	648	370	136	880	226.19	390
2014	3093	690	420	277	920	231.59	610
2015	3274	700	430	240	950	232.63	750
2016	3241	705	436	176	984	243	720
2017	3343	710	449	191	1020	253	910
2018	3330	748	462	292	1086	263	960
2019	3497	780	486	191	1135	249	1000

In addition, the index that affects the conditions of economic development is  $x_i = (i=1, 2, \dots, 7)$ , recorded as, then the shortest distance method is used to measure the distance between the class and the class among the index sample classes:

$$D(G_1, G_2) = \min_{\substack{x_i \in G_1 \\ y_j \in G_2}} \{d(x_i, y_j)\} \quad (2)$$

Intuitively indicate the closest distance between two classes to measure the distance of the sample class. Import the index data of Table 1 into the program to calculate the shortest distance between the final class and the class, and generate the distance matrix:

$$A = \begin{bmatrix} 0 & 2.2779 & 2.5557 & 2.7609 & 2.0267 & 2.7425 & 2.3391 \\ 2.2779 & 0 & 2.2778 & 0.4830 & 0.2512 & 0.4646 & 0.2006 \\ 2.5557 & 2.2778 & 0 & 0.2052 & 0.5290 & 0.1868 & 0.2454 \\ 2.7609 & 0.4830 & 0.2052 & 0 & 0.7342 & 0.0464 & 0.4218 \\ 2.0267 & 0.2512 & 0.5290 & 0.7342 & 0 & 0.7158 & 0.3124 \\ 2.7425 & 0.4646 & 0.1868 & 0.0464 & 0.7158 & 0 & 0.4034 \\ 2.3391 & 0.2006 & 0.2454 & 0.4218 & 0.3124 & 0.4034 & 0 \end{bmatrix}$$

## 3. CLUSTER ANALYSIS RESULT

Perform cluster analysis on the evaluation index system established above according to the index sample class shortest distance matrix, calculate the distance from each new sample point (unknown category) to historical sample point (known category), and then predict the category of the new sample as The category of the most similar historical sample point. Or predict the category of the new sample as the most similar among the multiple historical sample points. The basic idea of discrimination is to classify the samples with the closer distance into one class, and the samples with the larger distance into different classes.

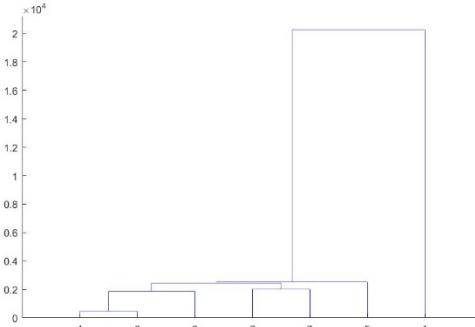


Figure 2 Cluster diagram of U.S. economic impact indicators system

According to Figure 2, it is appropriate to divide the set of indicators that affect US economic development into five categories. Among them, economic taxation, environment and national defense, and medical insurance are a single category, employment and immigration, and trade balance are in one category. Basic settings, Education investment is one category.

## 2.2 ECONOMIC DEVELOPMENT FORECAST BASED ON GM(1, 1)

Combining the above cluster analysis results of the seven indicator systems of taxation, infrastructure, environmental protection and national defense, employment and immigration, medical insurance, trade

balance, and education investment, the inductive analysis yields five key aspects that determine economic development, respectively[4]. *A.* Financial aspects (including taxation), *B.* environmental aspects (including environmental protection and national defense), *C.* external aspects (including employment and immigration and trade balance), *D.* people's livelihood aspects (including infrastructure and education) and *E.* medical aspects (including medical insurance).

In order to quantitatively describe economic changes and more accurately describe the economic impact of the election of different candidates in the 2020 U.S. election, gross domestic product (GDP) is used to calculate the final outcome of production activities of all resident units in the country during a certain period of time, as An indicator to measure the state of the country's economy. Obtain the quantitative relationship between economic development and key aspects:

Step1:Construct the original series of the total annual GDP data of the United States from 2017 to 2020 during Trump's administration,  $X^{(0)} = (19386.8, 20500.6, 21427.1, 21664)$ .

Step2:Find the level ratio  $\lambda^{(k)}$ , there is

$$\lambda(k) = \frac{x^{(0)}(k-1)}{x^{(0)}(k)} \quad (4)$$

Because of all,  $\lambda^{(k)} = [0.995, 1.045]$ , k=2, 3, 4 it can be used as a satisfactory GM(1, 1) model.

Step3:Accumulate the original data  $X^{(0)}$  once to get:

$$x^{(1)} = (19386.8, 39887.4, 61314.5, 82978.5)$$

Step4:Construct data matrix  $B$  and data vector  $Y$ , there are:

$$B = \begin{bmatrix} -\frac{1}{2}(x^{(1)}(1) + x^{(1)}(2)) \\ -\frac{1}{2}(x^{(1)}(2) + x^{(1)}(3)) \\ -\frac{1}{2}(x^{(1)}(3) + x^{(1)}(4)) \end{bmatrix}, Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ x^{(0)}(4) \end{bmatrix} \quad (5)$$

Calculation

$$\hat{u} = \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = (B^T B)^{-1} B^T Y = \begin{bmatrix} 0.59 \\ 0.09 \end{bmatrix} \quad (6)$$

So get:  $\hat{a} = 0.59$ ,  $\hat{b} = 0.09$ .

Step5:Modeling:

$$\frac{dx^{(1)}}{dt} + \hat{a}x^{(1)} = \hat{b} \quad (7)$$

Step6:Solve the forecast equation for the total amount of US GDP data under Trump's re-election:

$$X^{(1)}(k+1) = 3.808e^{0.04983t} - 2.294e^{1.983t} - 3.623 \times 10^5 \quad (8)$$

In the same way, the GM (1, 1) prediction model relational expressions of other five aspects that affect economic development are obtained:

$$\begin{cases} A.X^{(1)}(k+1) = 95899.0e^{0.03451t} - 0.0001518e^{3.547t} - 92544.0 \\ B.X^{(1)}(k+1) = 11200.0e^{0.04065t} - 0.02691e^{2.119t} - 10755.0 \\ C.X^{(1)}(k+1) = 12600.0e^{0.02057t} - 4.366e^{-7}e^{4.195t} - 12355.0 \\ D.X^{(1)}(k+1) = 13222.0e^{0.05558t} - 48.07e^{0.7432t} - 12466.0 \\ E.X^{(1)}(k+1) = 4627.0e^{0.009807t} \sin(0.1268t) - 3307.0e^{0.009807t} \cos(0.1268t) + 4327.0 \end{cases}$$

In addition, according to the GM (1, 1) model, during Trump's administration from 2017 to 2020, the US national fiscal revenue, environmental protection, development of new energy, national defense construction, external construction, and peoples livelihood infrastructure can be obtained from 2017 to 2020. Figure 3 shows the state's financial investment in medical treatment and its trend of change.

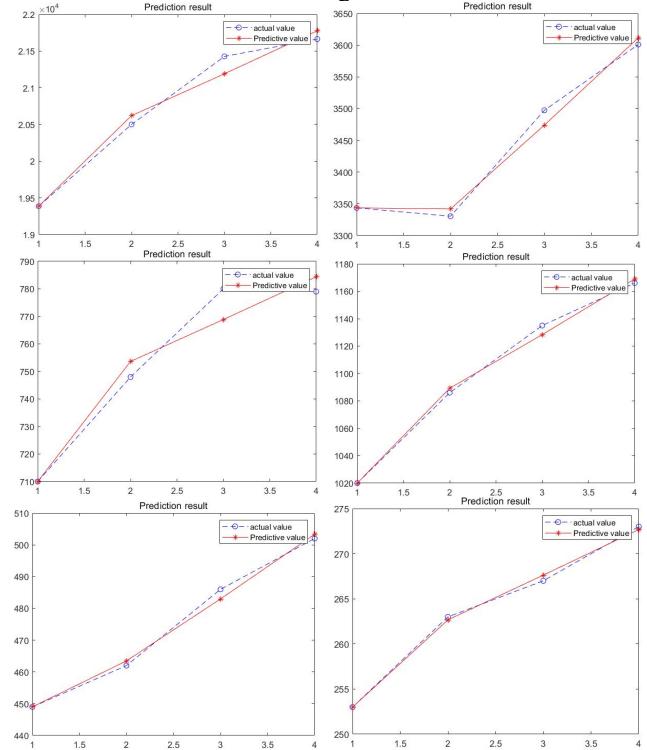


Figure 3 Economic development forecast for Trump's re-election in 2021

From top to bottom, from left to right, they are GDF, *A.*, *B.*, *C.*, *D.*, *E.* aspect.

It can be seen from Figure 3 that during Trump's administration, the capital investment in national construction and people's living infrastructure has shown an increasing trend year by year. In fact, during his tenure, the U.S. economy showed an upward trend year by year, and the social unemployment rate also broke a 50-year low in U.S. history, which is in line with actual conditions. Finally, the GM (1, 1) model predicts the situation of Trump's re-election in the 2020 U.S. election, corresponding to the state and society's construction capital investment in all aspects of the country and society in 2021 and the forecast data of U.S. GDP, as shown in Table 2. Shown.

Table 2 Economic Status of Trump's Re-election in 2021  
(Unit: USD 1 billion)

Project	<i>A.</i>	<i>B.</i>	<i>C.</i>	<i>D.</i>	<i>E.</i>	GDP
Capital investment	3699	521	274	792	1209	23086

### 2.3 U.S. ECONOMIC IMPACT BASED ON MULTIPLE LINEAR REGRESSION

Candidates' policy propositions in different fields have their own priorities. The two parties have certain differences in the formulation of economic blueprint

policy plans for the United States after taking office. Through their policy guidelines, they act on various factors affecting the economy, and quantitatively analyze the influence of different candidates on the United States.

### 1. TRUMP IS ELECTED:

Step1:Trump's tenure as President of the United States is roughly from 2017 to 2020. Only a 4-year period is not very accurate for multiple linear regression to a certain extent. In order to make the quantitative analysis of Trump's impact on the US economy after his re-election more accurate, First use the GM (1, 1) model to predict the annual GDP and various data of the United States from 2014 to 2016. If Trump has been in power for 7 consecutive years, the impact of his policy propositions and policy guidelines on the US economy will be more significant. as shown in Table 3.

Table 3 Forecast of economic data from 2014 to 2016

Year	GDP	A.	B.	C.
2014	17418.9	3093.4	420	231.59
2015	17937.8	3274.8	430	232.63
2016	18566.9	3241.8	436	243

Step2:Draw a scatter plot of the independent variables of fiscal, environmental, external, people's livelihood and medical aspects of the new crown pneumonia epidemic and the dependent variable of the US GDP to observe whether there is a linear relationship, as shown in Figure 4.

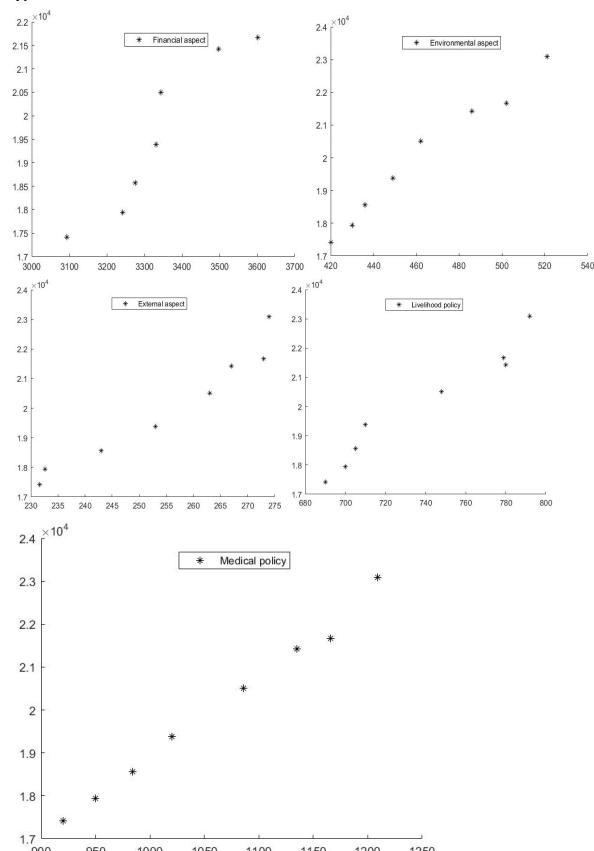


Figure 4 Scatter plot of dependent and independent variables

From top to bottom, from left to right, they are A. | B. | C. | D. | E. aspect. According to Figure 4, it can be obtained that the relationship between the independent variables

and the dependent variables is roughly linear.

Step3:The model of multiple linear regression analysis is:

$$\begin{cases} y = \beta_0 + \beta_1 x_1 + \beta_m x_m + \varepsilon \\ \varepsilon \sim N(0, \sigma^2) \end{cases} \quad (9)$$

Among them,  $\beta_0, \beta_1, \dots, \beta_m, \sigma^2$  are unknown parameters irrelevant to  $x_1, x_2, \dots, x_m$  ( $m = 1, 2, 3, 4, 5$ ). Among them,  $\beta_0, \beta_1, \dots, \beta_m$  is called the regression coefficient of the model.

Step4:Considering all the independent variables, the multiple linear regression equation is obtained:

$$y = 232.9 - 2.2A + 26.2B + 1.9C + 18.1D + 1.4E. \quad (10)$$

Step5:It shows that this multiple linear regression equation is highly significant, indicating that the overall independent variables of fiscal, environmental, external, people's livelihood and medical aspects of the new crown

pneumonia epidemic have a highly linear relationship with the dependent variable U.S. GDP, and the model fits well.

According to the above linear regression equation, if Trump is re-elected as president of the United States in 2020, it will have the greatest impact on the U.S. economy in terms of environmental protection, national defense construction, social infrastructure, education investment, etc. In addition, Trump's attitude towards polarization between the rich and the poor and class gap is also quite different. The implementation of the policy is to shape the dividend coverage through continuous tax reduction. The continued implementation of this policy may affect the development of national finance and taxation to a certain extent, resulting in its side effects on the sustained growth rate of the U.S. economy.

### 2.BIDEN IS ELECTED:

Two elections for president in 1988 and 2008 ended in failure. Although Biden had no previous experience in running for the president of the United States, he served as Vice President during Obama's term of office from November 2008 to January 2017 for an eight-year term[5]. In the 2020 U.S. election, predicting that Biden's impact on U.S. economic development after the election cannot be the same as the evaluation method of Trump's election, and there is no data during the president's administration. Therefore, the economic data during Obama's administration is used as the observation value, as shown in Table 4. Through multiple linear regression analysis, it studies the impact of Biden's campaign platform and policy propositions on the US economy in 2021 after his election.

Table 4 U.S. economic observation data from 2009 to 2016

Year	A.	B.	C.	D.	E.	GDP
2009	2044	306	124	432	702	13939
2010	2205	324	178	459	733	14526.5
2011	2326	347	159	501	769	15094
2012	2509.2	362	190	541	797	15684.8
2013	2882.3	375	226	578	838	16768.1
2014	3093.4	426	231	619	883	17418.9
2015	3274.8	430	232	655	935	17937.8
2016	3241.8	436	243	682	982	18566.9

Step1:Biden served as the Vice President of the United

States for eight years. In the same way as described above, draw a scatter plot of independent variables of fiscal, environmental, external, people's livelihood and medical aspects of the new crown pneumonia epidemic and the dependent variable of the US GDP to observe whether it is It has a linear relationship, as shown in Figure 5.

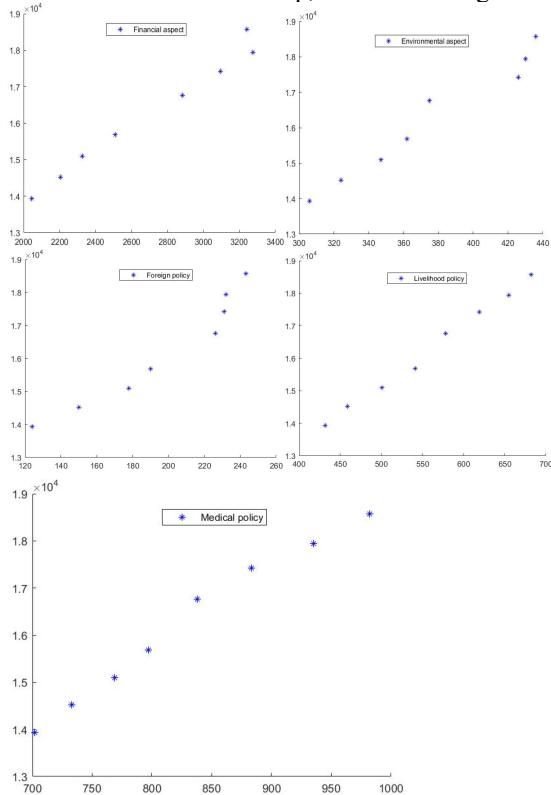


Figure 5 Scatter plot of dependent and independent variables

From top to bottom, from left to right, they are A. | B. | C. | D. | E. aspect. According to Figure 5, it can be obtained that the independent variable US economy and the dependent variables are roughly linear.

Step2: Considering all the independent variables, the multiple linear regression equation is obtained:

$$y = 5155.5 + A. + 2B. + 3.C. + 6.7D. + 5.8E. \quad (11)$$

Step3: Statistical analysis:  $F=213$ ,  $P=0.00004133 < 0.01$ ,  $R^2=1.04$  is very close to 1. It shows that this multiple linear regression equation is highly significant, indicating that the overall independent variables of fiscal, environmental, external, people's livelihood and medical aspects of the new crown pneumonia epidemic have a highly linear relationship with the dependent variable U.S. GDP, and the model fits well.

According to the above linear regression equation, if Biden is elected as the President of the United States in the 2020 U.S. election, his policies in social infrastructure, education investment, and medical care will have the greatest impact on the U.S. economy. As we all know, Biden and Trump have the greatest influence on the new president. The virus epidemic has very different attitudes towards air defense and vaccines. At the same time, the government's fiscal revenue has remained stable to a certain extent, which is conducive to the sustained and

stable growth of the US economy.

### 3. CONCLUSIONS

Trump was re-elected as President of the United States, and Biden was elected President of the United States. The total amount of US economic data and linear prediction results in 2021 are shown in Table 5.

Table 5 Comparison of Economic Data Volume (Unit: USD 1 billion)

Projec t	A.	B.	C.	D.	E.	GM(1, 1)GD P	Linear GDP
Trump	369	521	274	792	120	23086	22294.0
Biden	360	502	273	779	116	-	22752.7

According to Table 5, after Trump's re-election, the US GDP in 2021 is 22, 294, and the GM (1, 1) result is 23, 086, indicating that Trump's re-election may lead to US economic output if the current political propositions continue. The decline is not conducive to the development of the U.S. economy. After Biden's election, the U.S. GDP in 2021 is 22752.7, and the result obtained through the multiple linear regression model is 21644. The international economic structure has given the US economic aggregate a certain increase, which is conducive to the healthy and sustainable development of the US economy.

Biden and the Democrats have been pushing for a new round of economic stimulus since a new crown pneumonia epidemic hit in 2020, and plan to appoint Fed doves Brainard as Treasury secretary. Biden will continue in his new role next year. The US is expected to continue to release liquidity to hedge against the impact of the pandemic, which will help the us recovery to some extent. Moreover, emerging markets have the comparative advantage of higher yields against a background of zero or even negative funding rates in developed markets. If Mr Biden wins the US presidential election in 2020, a flood of money is likely to pour into emerging markets, according to bank analysis. Before the new coronavirus epidemic, tariffs and Chinese procurement commitments slightly reduced the U.S. trade deficit with China, but the epidemic fully reversed that deficit.

Politics, especially political leaders, are the chief helmsman of the economy. If Trump is re-elected in the U.S. election in 2020, his future will be obsessed with economic liberalism. The U.S. economy can still evolve smoothly in the capital market; if Biden is elected, the New Deal liberalism he advocates can be after Obama. A second sprint can gain ground, and it will also lead to a more open trade policy[6]. It may improve the US trade environment in the world structure, but its open and free attitude towards Wall Street will cause more uncontrollable capital markets.

### REFERENCES

- [1] Diao Daming. The particularity and impact of the 2020 U.S. election [J]. Modern International Relations 2020(08): 9-16+61.
- [2] Zhang Rui. The presidential election affects future changes in the US economy [N]. International Finance

News, 2020-03-16 (003).

[3] Kuang Tao. The role of the US vice president in US diplomatic decision-making[J]. International Research Reference, 2015(09): 9-15.

[4] Zhang Rui. The presidential election affects future changes in the US economy [N]. International Finance News, 2020-03-16(003).

[5] Edmund Phelps. The impact of the long-term

economic slowdown in the United States, the loss of investment "spirit", and the substantial increase in public debt on China [J]. China Economic Report, 2020(05): 121-126.

[6] Hong Lu. Who is optimistic about emerging markets to be elected president of the United States. International Business Daily, 2020-11-13(004).

# Food System Committed to Fair and Sustainable Development

Xichang Li<sup>1, 2, \*</sup>, Shihao Lu<sup>1, 3</sup>, Kunlong Li<sup>1, 4</sup>

<sup>1</sup>Mathematical Modeling Innovation Lab North China University of Science and Technology, Tangshan 063210, China;

<sup>2</sup>School of chemical engineering, North China University of Science and Technology, Tangshan 063210, China;

<sup>3</sup>School of Artificial Intelligence, North China University of Science and Technology, Tangshan 063210, China;

<sup>4</sup>School of science, North China University of Science and Technology, Tangshan 063210, China

\*Corresponding Author.

**Abstract:** Food security issues identified in the Global Food Crisis Report will create new risks for some countries. In order to implement the concept of sustainable development and reduce food waste, this paper constructed a food system evaluation model based on time series model. The model by using the principal component analysis, KMO and Bartlett sphericity test for data preprocessing, after will get the data using the time series models to forecast, finally compared with the actual and estimated values, found the model for China, the United States and Algeria regions such as the food system evaluation has good applicability.

**Key words:** Principal Component Analysis; Food Security; Time Series Model

## 1. BACKGROUND

The disasters of the past year have shown the fragility of the human food system and the suffering it inflicts on the world. Modern food systems are more efficient and profitable, but even with enough food to feed everyone, more than 800 million people around the world go hungry. On top of that, the current food system has a huge negative impact on biodiversity, accounts for a large proportion of freshwater use, and has a huge impact on the natural environment in which we live. Therefore, this paper constructs a food system evaluation model based on time series model.

## 2. FOOD SYSTEM EVALUATION

Efficiency and profitability are the mainstream mode of business today, but environmental damage is becoming more and more serious as the price of rapid economic development. So sustainable development should begin to become more valued.

According to surveys, rice is the staple food of about 50% of the world's population, and wheat is also one of the staple foods established by the United Nations, which can produce flour, which can then be made into various products such as bread, steamed bread, noodles and so on. In addition, the world's wheat has a large planting span and a long vertical, mainly distributed below 3000 meters above sea level. In order to simplify the model, we mainly choose agricultural crops such as wheat and rice as representatives of the food system. By analyzing some data from the three countries of the United States, China and Algeria, we have established a food system evaluation model.

### 2.1 CURRENT PARTIAL FOOD SYSTEM

Food waste will generate huge carbon dioxide emissions, which is equivalent to a large-scale and rapid industrial development of greenhouse gas (GHG) emitters. We know that the amount of carbon dioxide produced by the food loss caused by the not advanced and not green agricultural production is much lower than the amount of carbon dioxide produced by food waste at the consumer level<sup>[1]</sup>. This indicates that changes in consumer behavior may have a greater impact on reducing greenhouse gas emissions, thereby improving the fairness and sustainability of the food system.

## 3.NATIONAL INITIATIVES

And we know that the laws of China, the United States and other major countries do not enact severe penalties for the waste of food. At most, they only discourage the waste or encourage the people to save food. According to the survey, in real life, more people throw away food as simple as throwing away garbage. But what we should not ignore is that many people in the world are still starving. In the face of massive food waste, many countries have also adopted their own measures. For example, The US Food and Drug Administration requires companies to use the "best before date" to standardize the expiration date to reduce food waste; France stipulates that supermarkets must not discard or destroy unsold food, but can donate it to charities and food bank. Japan began to introduce the concept of a food bank around 2000. Food banks usually collect safe and edible food from retailers and distribute them to those in need<sup>[2]</sup>.

### 3.1 OUR IDEA-TO RESIST DISASTERS TO BE SELF-RELIANT

The current food system is restricted by different laws due to different countries, but it still cannot stop the occurrence of food waste, making food insecurity exist all over the world. Through the survey data<sup>[3]</sup>, we believe that the behaviors that can guarantee food security are mainly to prevent natural disasters and develop agricultural technology. In addition, the most basic way to ensure food security is to be self-sufficient, so we suggest that different regions adopt different treatment methods to ensure food production and sustainable development.

Therefore, we suggest building water conservancy projects in flooded areas, locust-prone locusts, actively controlling locusts in advance, developing drip irrigation technology in dry areas, developing greenhouse economy in cold but fertile soil areas, and areas lacking education should focus on the local government.

#### 4. MODEL SOLVING

##### 4.1 CALCULATION OF INDICATORS

We use SPSS to perform principal component analysis on various indicators of efficiency, profitability, fairness, and sustainability. First, import the data into SPSS and select the factor analysis in the dimensionality reduction analysis. Import the various indicators of efficiency, profitability, fairness, and sustainability into the variables, and then select the description, and select the univariate description and initial solution in the statistics and the coefficients in the correlation matrix, the significance level, KMO and Bartlett Check the sphericity, and then select the correlation matrix, the unrotated factor solution and the gravel map in extraction After running, it is found that two principal components will be extracted, so return to the extraction and select the fixed number of factors to be extracted as 2. After running, the KMO and Bartlett sphericity test and the gravel map are as follows:

Table 1. KMO and Bartlett test1

KMO and Bartlett test	
KMO Sampling suitability	.635
Bartlett sphericity test	
Approximate chi-square	57.126
Degree of freedom	10
Significance	.000

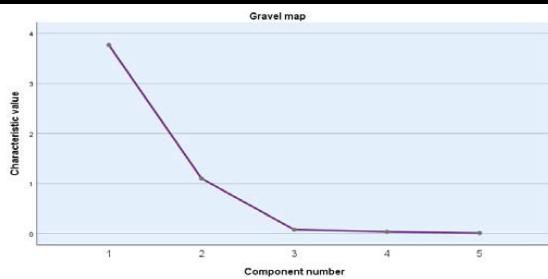


Figure 1: Component number first

Table 2. KMO and Bartlett test2

KMO and Bartlett test	
KMO Sampling suitability	.461
Bartlett sphericity test	
Approximate chi-square	35.754
Degree of freedom	10
Significance	.000

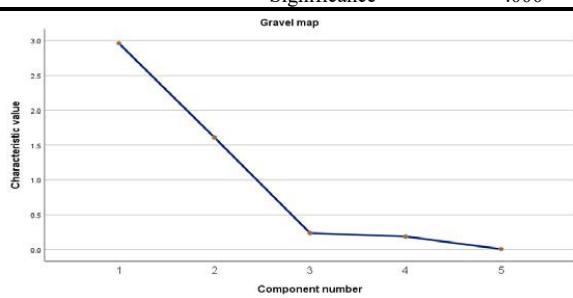


Figure 2: Component number second

According to the SPSS operation, the principal components obtained after the efficiency and profit operation are the same, the principal components obtained after the fair and sustainable operation are the same, and according to the KMO and Bartlett sphericity test, the significance of the chi-square test is less than 0.5, Reject the null hypothesis, so factor analysis can be used. According to the gravel chart, two principal components can be extracted, so in this question we choose the output of wheat as the main indicator of efficiency, the import

price of wheat as the main indicator of profitability, and the producer price of wheat as the main indicator of fairness. The CO2 emissions are the main indicators of sustainability.

##### 4.2 IMPLEMENTATION TIME

For the time when the system can be implemented, we use the time series model in SPSS to predict the CO2 emissions of rice from 2020 to 2030 and the producer price of wheat. First, we will calculate the CO2 emissions of rice and the production of wheat from 2010 to 2019. The price of each is imported into SPSS, Position the time in 2010, then choose to create a traditional model in the time series forecast, import the CO2 emissions of rice and the producer price of wheat into the dependent variable, and import the time into the independent variable. The conditional model selects all models, Select additive, level change, innovation, transient, local trend, additive repair for outliers, select the parameter estimates in the statistics of a single model in the statistics, select the sequence of the single model in the figure, and the measured value, Predicted value, fitted value, confidence interval of predicted value, confidence interval of fitted value, residual autocorrelation coefficient, residual partial autocorrelation coefficient, Then select the saved forecast, and select 11 years of forecast in the options<sup>[4]</sup>. After running, if each forecast value is the same, return to delete the time in the independent variable, select the exponential smoothing model, run again, and finally get the forecast result as shown in the figure Show:

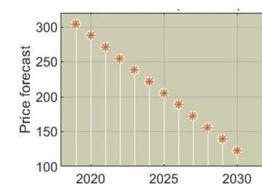


Figure 3: Forecast of wheat producer price

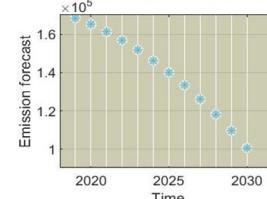


Figure 4: Prediction of CO2 emission from Rice

According to the figure, if the food system is optimized for fair and sustainable development, the producer price of wheat and the CO2 emissions of rice will continue to decline. According to national policies, it is estimated that when the producer price of wheat is between 100-150 US dollars, It will be fairer, that is, the number of hungry people will decrease. According to the picture 3, it can be seen that the system may be realized in 2029; At the same time, when the CO2 emission of rice is about 160,000 or less, the emission standard will be met, that is, the pollution degree may have a degree of decline. According to picture 4, it can be seen that the system may be realized when the time is 2029. As mentioned above, if the food system is optimized for fair and sustainable development, the system may be realized in 2029.

##### 4.3 Entropy Method To Measure Weight

The steps of entropy method to calculate weight are as follows:

step 1: Because there is a big difference between the data values of the indicators and the indicators, we first normalize each indicator to determine whether the indicator is a positive indicator or a negative indicator.

For example, we hope that the larger the crop yield, the better, and the crop yield is a positive indicator. It is hoped that the smaller the carbon dioxide emissions, the better, and the carbon dioxide emissions are a negative indicator.

$$x'_{ij} = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)} \text{ Or } W_j = (1 - e_j)/\sum(1 - e_j), j = 1, 2, \dots, n$$

The first formula is a positive indicator, and the second formula is a negative indicator.  $x_{ij}$  represents the i-th evaluated value under the j-th index under the food system,  $\min(x_j)$  represents the minimum value under the j-th index under the food system, and  $\max(x_j)$  represents the maximum value under the j-th index under the food system value.

Step 2: Calculate the proportion of each indicator

$$p_{ij} = \frac{x_{ij}}{\sum_i x_{jj}}$$

Among them,  $p_{ij}$  represents the proportion of the i-th evaluated value under the j-th index in the food system to the index.

Step 3: Calculate the entropy value of each indicator

$$e_j = -k \sum_{p_{ij}} \ln p_{ij}, k = 1/\ln m$$

Where  $e_j$  represents the entropy value of the j-th index under the food system, and k is a constant.

Step 4: Calculate the weight of each indicator

$$W_j = (1 - e_j)/\sum(1 - e_j), j = 1, 2, \dots, n$$

$W_{ij}$  is the weight of each indicator under the food system. We use matlab to calculate efficiency, profitability, fairness and sustainability of the various indicators in the food system to account for the weights, first import the data into matlab, write the code according to the above steps, the weight results are as follows:

Table 3. Weight result

	Efficiency	profitability	fairness	sustainability
index	Wheat yield	Wheat import prices	Producer price	CO2 emissions from rice
W	0.1955	0.4409	0.1391	0.2243

#### 4.4 ANALYTIC HIERARCHY PROCESS TO MEASURE WEIGHT

The process of analytic hierarchy process to obtain index weight is as follows:

Step 1 Analyze the research problem, determine the overall goal of the system, collect the criteria that affect the decision-making of the goal, and use efficiency, profitability, sustainability, and fairness as the criterion layer.

Step 2 Establish a hierarchical structure model, divide each element related to decision-making into different levels, and upper-level elements are used as guidelines to dominate lower-level elements.

Step 3 Construct a pair of judgment matrices to determine the relative importance of each element in a certain layer to the upper-layer element, which is expressed by the relative scale  $a_{ij}$ .

Step 4 Calculate the single ranking weight vector and perform consistency check. If the test passes, the feature vector (after normalization) is the weight vector; if it fails,

it needs to be reconstructed into a pair comparison matrix. The maximum eigenvalue solved by matlab2017 is 4.0260, and the consistency test is performed. The random consistency index is 0.0087, which is less than 0.1, so we think it passed the test, and then solve the eigenvector of the maximum eigenvalue and normalize The weight vector of each indicator is  $W=(0.0970, 0.1637, 0.4020, 0.3372)$ .

The steps to check the consistency of the judgment matrix are as follows:

- Calculate the consistency index CI
- Find the corresponding average random consistency index RI. For  $n = 1, \dots, 9$ , Saaty gives the value of RI, as shown in Table 2:

Table 4. The value of RI

n	1	2	3	4	5	6	7	8	9
RI	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45

- The value of RI is obtained by using a random method to construct 500 sample matrices: randomly select numbers from 1 to 9 and its reciprocal to construct a reciprocal matrix, find the average value of the largest eigenvalue, and define

$$RI = \frac{\lambda' - n}{n - 1}$$

Calculate the consistency ratio CR:

When  $CR < 0.10$ , the consistency of the judgment matrix is considered acceptable, otherwise the judgment matrix should be appropriately modified.

$$CR = \frac{CI}{RI}$$

Among them are the CI consistency index, RI random consistency index, and CR consistency ratio.

In order to make up for the shortage of arable land resources, a large number of elements such as fertilizers and pesticides are used in food production to replace arable land resources. Although this has alleviated the contradiction between grain and cultivated land, it has aggravated the depletion of resources and ecological damage caused by grain production.

The weight comparison before and after optimization is shown in the figure:

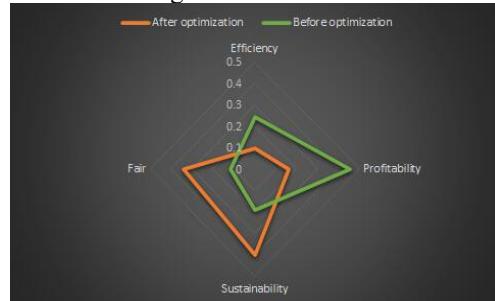


Figure 5. Weight comparison

#### 5.FOOD SECURITY

Countries around the world have made many efforts to ensure food security, but the global food security situation is still very serious. The current food system is dominated by efficiency and profitability, and the four staple foods established by the United Nations are all grown on the ground, so we will discuss the changes in the food system in terms of changes in agricultural production.

Change the weight, Consumers will have an impact: it will reduce food waste caused by consumers due to different economic levels. In order to ensure fairness, the government should reduce the problem of food insecurity caused by income disparity, and avoid the occurrence of fiscal deficits that make food price control impossible<sup>[5]</sup>.

### 5.1 THE PRICE PAY

**Cost:** We use the time series model to predict wheat production and wheat import prices as shown below:

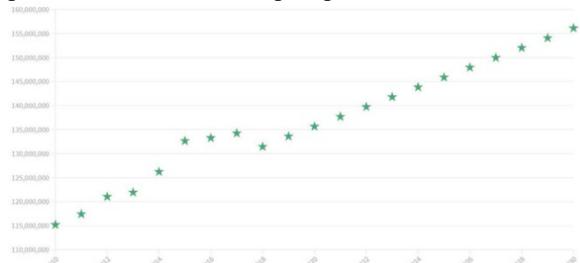


Figure 6. Prediction of wheat production

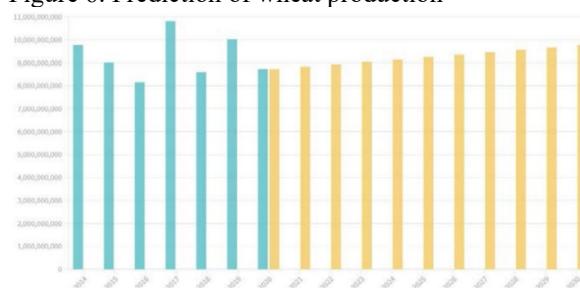


Figure 7. Prediction of wheat import prices

It can be seen from the figure that the overall economic growth trend before 2021 is an upward trend, but it is extremely unstable. After optimization, the entire curve is smoother, but the overall growth rate has become slower. That is to say, although the optimized food system can meet the country's food needs under normal conditions and its growth trend has become more stable, its growth rate is lower than the current development speed under the food system. It takes about ten years to complete the overall economic growth.

In addition, this policy has certain flaws. For China, the current food system is sufficient to provide sufficient food when some small natural disasters occur. But when an accident happens, no matter how big or small the accident is, it will always cause panic among many people, leading to panic incidents such as snapping up food and salt. For the time when the system occurs, we use the time series model in SPSS to predict the import price of wheat and the output of wheat from 2020 to 2030. and finally get the predicted result as shown in the figure below:

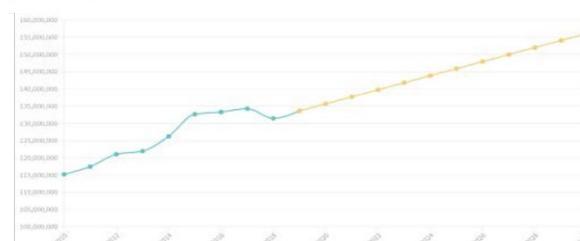


Figure 8.The prediction of the output of wheat from 2020 to 2030

According to the figure, if the priority supply order of the food system is changed, the import price of wheat and the growth of wheat production will slow down. According to the country's policies over the years, the efficiency of wheat production at around 150 million is low. According to Figure 8, it can be seen that it may be realized in 2028; for the same reason, according to the import price of wheat over the years, when the import price of wheat is around 950, the profit will be lower, that is, when the profit will be lower in 2027 or 2028, and According to question 1, the CO<sub>2</sub> emissions of rice and the producer price of wheat will reach a relatively satisfactory level in 2029. In summary, the priority order of supply in the food system will be changed from 2028 to 2029. The benefits and costs may occur.

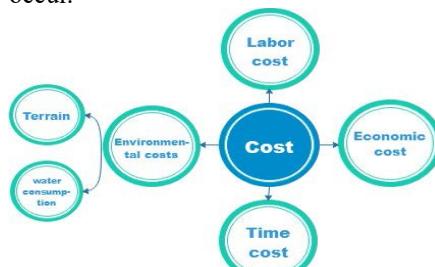


Figure 9. Influencing factors of cost

### 6. SENSITIVITY ANALYSIS

The data is obtained through statistical observation, so we need to consider the possibility of inaccurate data. On the basis of the food system evaluation model, we first select the three indicators of food production, producer price, and CO<sub>2</sub> emissions to represent profitability, and fairness and sustainability as variables for local sensitivity analysis. In the analysis process, only the CO<sub>2</sub> emissions are changed under the condition that the grain output and the producer price remain unchanged, and the sensitivity of the model is measured by calculating the change in the output value of the model when the parameters change. The parameters are slightly disturbed during the calculation, such as a 5% change, the response fluctuation of the model output to a single input is the sensitivity index. The sensitivity analysis result is shown in the following figure:

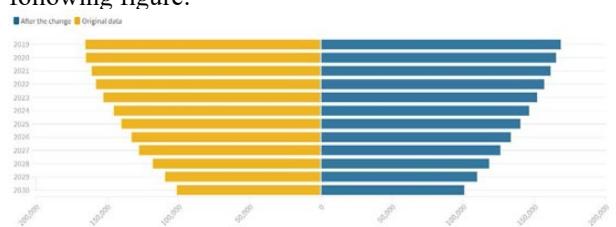


Figure 10. The result of sensitivity analysis

As can be seen from the figure, the trend of each color in the figure has not changed much, and the general trend after the change is the same as the data trend before the change. In summary, the sensitivity is superior.

### 7. CONCLUSION

Model evaluation:

In the process of solving the problem, we used the principal component analysis method, time series forecasting model, analytic hierarchy process and entropy

weight method.

Advantages:

1.The principal component analysis method can eliminate the relative influence between evaluation indexes and reduce the workload of index selection.

2.In the analytic hierarchy process, the degree of influence of each factor in each level on the results is quantified and very clear.

3.The data used are all real-time updated data with high accuracy.

Disadvantages:

1.The interpretation of principal components generally has a somewhat vague meaning, not as clear and precise as the meaning of the original variables.

2.The time series forecasting method does not consider the influence of external factors, so it has the defect of forecast error and is not suitable for long-term forecasting.

3.The analytic hierarchy process has less quantitative data and many qualitative components, which is not easily convincing.

The model has limitations in its adaptability in some special countries.

## REFERENCES

- [1] Boliko Mbali Charles. FAO and the Situation of Food Security and Nutrition in the World [J]. *Journal of nutritional science and vitaminology*, 2019, 65(Supplement).
- [2] New laws on food conservation in various countries, retrieved from: [http://www.360doc.com/content/20/0920/08/4616181\\_936663279.shtml](http://www.360doc.com/content/20/0920/08/4616181_936663279.shtml).
- [3] Data collected from FAO(united nations food agriculture organization, FAO) <http://www.fao.org/statistics/zh/>.
- [4] Fu Gonghua, Yan Bangyou. Research on China's Future Food Security and Ecological Sustainability--Based on the Empirical Analysis of the Ecological Footprint of Food Production Process[J]. *Resources and Environment in the Yangtze River Basin*, 2013, 22(12):1550-1556.
- [5] Zhang Yuanhong, Liu Changquan, GuoLulai. Evaluation of China's Food Security Status and Strategic Thinking[J]. *China Rural Observation*, 2015(01).

# Evaluation of Mental Health Status of Minors Based on Improved Analytic Hierarchy Process

Kun Liu<sup>1</sup>, Shuo Zhang<sup>2</sup>, Xichang Li<sup>1</sup>

<sup>1</sup> School of Chemical Engineering, North China University of Science and Technology, Tangshan 063210, China;

<sup>2</sup> School of Basic Medical Sciences, North China University of Science and Technology, Tangshan 063210, China

\*Corresponding Author.

**Abstract:** According to statistics, at present, the number of minors with psychological problems has increased sharply across the country, which has become an important public health problem related to the future of the country and the nation. In this paper, we establish a mathematical model of minors' mental health evaluation system to better understand minors' mental health. Firstly, we issued a questionnaire and collected 1500 mental health data of middle school students in Tangshan. We selected 11 items in the data, such as gender, obsessive-compulsive symptoms, paranoia, hostility, interpersonal tension and sensitivity, depression, anxiety, learning pressure, maladjustment, emotional imbalance and psychological imbalance, as evaluation indicators, and then combined with K-means clustering to cluster these 11 indicators into five categories. Finally, the improved analytic hierarchy process is used to calculate the relative weight of each index, sort the obtained weight, delete the four indexes with the smallest weight, and update the weight of the remaining seven indexes that have a great impact on the mental health of minors. Respectively  $P_1 = 0.4380$ ,  $P_5 = 0.0240$ ,  $P_6 = 0.0406$ ,  $P_8 = 0.0252$ ,  $P_9 = 0.2449$ ,  $P_{10} = 0.1136$ ,  $P_{11} = 0.1136$ . Establish the mental health evaluation system model of minors. Among them, we assume that the scores of the minors surveyed can truly reflect their psychological status, and that their environment is not a major disaster.

**Keywords:** K-Means Clustering; Improved Analytic Hierarchy; Process Mental Health

## 1.BACKGROUND

Today, we are in an era of rapid development and volatility. In daily life, minors may be due to their own immaturity, or received the influence of family, society and other external environment, in the psychological formation of unhealthy factors. Teenagers are often mentally immature and unable to deal with problems rationally on their own, which leads to their inability to know how to solve psychological problems when they arise. If society, school and family do not give correct and reasonable guidance to minors, their mental health may be affected. Therefore, we must strengthen the attention to the mental health of minors, to eliminate their psychological problems to the greatest extent, it is particularly important to build a reasonable evaluation system for the mental health of minors.

## 2.K-MEANS CLUSTERING TO DETERMINE INDEX CLASSIFICATION

We use this formula to calculate the Euclidean distance

from each object to each cluster center:

$$dis(X_i, C_j) = \sqrt{\sum_{t=1}^m (X_{it} - C_{jt})^2}$$

Where X represents the data sample containing n objects, where each object has attributes of M dimensions.

Next, we will use SPSS software to K-means[1] cluster the data: first, standardize the data, and then use the K-means[2] clustering function in classification to import all data into variables. The number of clusters is 5 and the maximum number of iterations is set to 10 to obtain the number of cases in each cluster, as shown in table 1. [3] According to table 1 and the operation results of SPSS, we divide the variables into five categories.

Table 1 number of cases in each cluster

Number of cases in each cluster		
	1	4.000
2		1.000
3		3.000
4		2.000
5		1.000
<i>Effective</i>		11.000
<i>defect</i>		.000

## 3. IMPROVING ANALYTIC HIERARCHY PROCESS AND ESTABLISHING EVALUATION SYSTEM

### 3.1 ESTABLISH A HIERARCHY MODEL:

Combined with the above K-means clustering results, we established a hierarchical model[4], as shown in Figure 2:

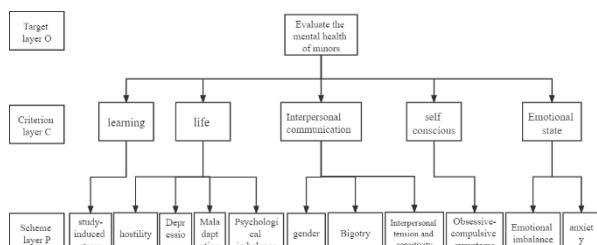


Figure 1 hierarchical structure model for evaluating the mental health status of minors

The hierarchy model is divided into three layers: target layer O, criterion layer C and scheme layer P. Among them, the target level O is the highest level, that is, to evaluate the mental health status of minors; Criterion layer C is the middle layer, Including learning  $C_1$ , life  $C_2$ , interpersonal communication  $C_3$ , self-consciousness  $C_4$ , emotional state  $C_5$  these five indicators; scheme layer P is the lowest layer, including learning pressure  $P_1$ , hostile  $P_2$ , depressed  $P_3$ , maladjustment  $P_4$ , psychological imbalance  $P_5$ , gender  $P_6$ , paranoid  $P_7$ , interpersonal tension and sensitivity  $P_8$ , obsessive compulsive symptoms  $P_9$ ,

emotional imbalance  $P_{10}$ , anxious  $P_{11}$  these 11 indicators.

### 3.2 CONSTRUCT JUDGMENT MATRIX[6]

Firstly, we construct the O-C three scale judgment matrix: here, we improve the traditional analytic hierarchy process. The traditional analytic hierarchy process uses the 1-9 scale to construct the judgment matrix, and this time we will use the 0, 1 and 2 scales to construct the judgment matrix, which will reduce the operation difficulty and make the analytic hierarchy process known for its strong subjectivity more objective. The O-C three scale judgment matrix is obtained by comparing the five indexes of  $C_1, C_2, C_3, C_4, C_5$  as shown in the table below:

Table 2 O-C three scale judgment matrix

$O$	$C_1$	$C_2$	$C_3$	$C_4$	$C_5$
$C_1$	1.0000	2.0000	2.0000	1.0000	1.0000
$C_2$	0.0000	1.0000	2.0000	0.0000	1.0000
$C_3$	0.0000	0.0000	1.0000	1.0000	0.0000
$C_4$	1.0000	0.0000	1.0000	1.0000	1.0000
$C_5$	1.0000	1.0000	2.0000	1.0000	1.0000

Then, we construct a  $C_x - P$  ( $x = 2, 3, 5$ ) three scale judgment matrix:

Table 3  $C_2 - P$  three scale judgment matrix

$C_5 - P$	$P_2$	$P_3$	$P_4$	$P_5$
$P_2$	1.0000	1.0000	1.0000	1.0000
$P_3$	1.0000	1.0000	1.0000	1.0000
$P_4$	1.0000	1.0000	1.0000	1.0000
$P_5$	1.0000	1.0000	1.0000	1.0000

Table 6 average random consistency index RI

$n$	1	2	3	4	5	6	7	8	9	10	11	12	13
RI	0	0	0.52	0.89	1.12	1.26	1.36	1.41	1.46	1.49	1.52	1.54	1.56

Note: when  $n = 1$  or  $n = 2$ , the judgment matrix is the consistency matrix, so  $CR$  is 0.

About  $C_2 - P$  ( $x=2, 3, 5$ ): We use the same method to obtain the relative weight vector and the maximum eigenvalue of the three judgment matrices  $\lambda_j$ , consistency index  $CI_j$ , consistency ratio  $CR_j$  ( $j=2, 3, 5$ ).

Table 7  $C_2 - P$  relative weight and consistency judgment index

$P_2$	$P_3$	$P_4$	$P_5$	$\lambda_2$	$CI_2$	$CR_2$
0.2296	0.2033	0.2631	0.3040	4.0000	0.0000	0.0000

Table 8  $C_3 - P$  relative weight and consistency judgment index

$P_6$	$P_7$	$P_8$	$\lambda_3$	$CI_3$	$CR_3$
0.4595	0.2558	0.2848	3.0000	0.0000	0.0000

Table 9  $C_5 - P$  relative weight and consistency judgment index

$P_{10}$	$P_{11}$	$\lambda_5$	$CI_5$	$CR_5$
0.5000	0.5000	2.0000	0.0000	0.0000

From the values in tables 7, 8 and 9,  $C_2 - P, C_3 - P, C_5 - P$  The three scale judgment matrices passed the consistency test.[8]

### 3.4 HIERARCHICAL TOTAL RANKING AND ITS CONSISTENCY TEST:

According to formula (4), combined with the relative weights of  $C_1, C_2, C_3, C_4, C_5$ , the consistency index  $CI$  and average random consistency index  $RI$  of C-P judgment matrix obtain the consistency ratio of total ranking,  $= 0 < 0.1$ , which passes the consistency test.

$$CR = \frac{a_1 CI_1 + a_2 CI_2 + \dots + a_m CI_m}{a_1 RI_1 + a_2 RI_2 + \dots + a_m RI_m} \quad (4)$$

Table 4  $C_3 - P$  three scale judgment matrix

$C_3$	$P_6$	$P_7$	$P_8$
$P_6$	1.0000	2.0000	0.0000
$P_7$	0.0000	1.0000	2.0000
$P_8$	2.0000	0.0000	1.0000

Table 5  $C_5 - P$  three scale judgment matrix

$C_5$	$P_{10}$	$P_{11}$
$P_{10}$	1.0000	1.0000
$P_{11}$	1.0000	1.0000

### 3.3 HIERARCHICAL SINGLE RANKING AND ITS CONSISTENCY TEST

For O-C: firstly, we use the eig () function of MATLAB software to obtain the maximum eigenvalue of O-C three-scale judgment matrix  $\lambda_1$  and the maximum eigenvector corresponding to maximum eigenvalue  $V_1$  (i.e. relative weight vector of 5 indicators),  $\lambda_1 = 3.99$ ,  $V_1 = (0.4065, 0.0734, 0.0820, 0.2273, 0.2108)^T$ , Then, the consistency index  $CI$  is -0.2536 obtained, and then the average random consistency index  $RI$  is 1.12. Then, the consistency ratio  $Cr$  is calculated,  $Cr = -0.0871 < 0$ , which passes the consistency test.[7]

The formula used is as follows.

$$CI = \frac{\lambda_m - n}{n - 1}$$

$$CR = \frac{CI}{RI}$$

Evaluating indicator  $P_1 \sim P_{11}$  the relative weights of are 0.4065, 0.0169, 0.0149, 0.0193, 0.0223, 0.0377, 0.0210, 0.0234, 0.2273, 0.1054, 0.1054. We ranked the relative weights of these 11 indicators from large to small, and found that the relative weights of  $P_2, P_3, P_4, P_7$  were very small, so we eliminated these four indicators to simplify our evaluation system. Finally, the relative weights of the seven indicators are shown in the table below:

Table 10 relative weights of seven indicators

Index	$P_1$	$P_5$	$P_6$	$P_8$	$P_9$	$P_{10}$	$P_{11}$
Relative weight	0.43 80	0.02 40	0.04 06	0.02 52	0.24 49	0.11 36	0.11 36

The final mental health evaluation system for minors is shown in Figure3:

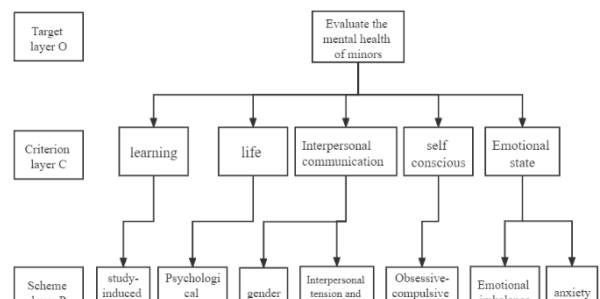


Figure 2 mental health evaluation system of minors

### 4.TO ANALYZE THE MENTAL HEALTH STATUS OF MINORS IN TANGSHAN

For the analysis of the mental health of minors in our

region, we need to analysis the region of each surveyed the mental health of minors, which calculated the area of each of the mental health of minors evaluation index, and then sum, average, get the area of the mental health of minors evaluation indexes. As a representative of the mental health of all minors in the area.

In order to evaluate the mental health status of every minor in this area, we established an evaluation model for the mental health status of minors.

Based on TOPSIS method and combined with improved analytic hierarchy process, the relative weight of 7 indicators is obtained, and the mental health score of minors to be evaluated is calculated by using the data of 7 indicators. The specific process is as follows[9]:

Firstly, the data of 7 indicators in the data were extracted as the pending data of TOPSIS, and then they are forward normalized and standardized. The maximum value of each indicator is the positive ideal solution, and the minimum value is the negative ideal solution, the positive distance of each juvenile is calculated by combining the relative weights of 11 indicators. The negative distance of each juvenile is calculated by combining the relative weights of 11 indicators, The formula used is as follows. Finally, the evaluation index value of each juvenile's mental health is obtained. [11]

$$\bar{x}_i = \frac{1}{x_i}, \quad z_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^n x_{ij}^2}}$$

$$D_i^+ = \sqrt{\sum_{j=1}^m W_j (Z_j^+ - z_{ij})^2}, \quad D_i^- = \sqrt{\sum_{j=1}^m W_j (Z_j^- - z_{ij})^2}$$

$$S_i = \frac{D_i^-}{D_i^+ + D_i^-}$$

Then, we use the mean() function of the array of MATLAB software to obtain the mean value of the mental health evaluation index of 1456 minors, which is 0.5582, as the representative of the mental health status of all minors in this area.

Through the analysis of the final score of  $0.5582 < 0.6$ , we believe that the mental health status of the minors in the region as a whole did not achieve the pass mark, so we came to the conclusion that the mental health of minors in the region, there are some problems, we should pay attention to the region's schools and students' parents, pay more attention to the mental health of minors, Remove their psychological barriers, so that minors can better grow up.

## 5. MODEL TEST

(1) Sensitivity analysis of mental health status assessment model

Here, we will conduct sensitivity analysis on the assessment model of juvenile mental health status composed of improved analytic Hierarchy Process +TOPSIS. Through literature review, we know that in the process of establishing the evaluation model, the weight of indicators plays a very important role in the evaluation result, so we will adjust the index with the maximum weight among the 7 indicators in the model.

First of all, we adjusted the weight of the index from 43.80% to 45.00%, while the weight of the remaining 6 indexes

was reduced by 0.20%. Then, we used the adjusted weight and the pre-processed data of 7 indicators, combined with TOPSIS method, to evaluate the mental health status of minors, summing up and averaging, and the result was reduced by 3.21%.

Then, we adjusted the weight of the index from 43.80% to 42.00%, and increased the weight of the other 6 indexes by 0.30%. Then, we used the adjusted weight and the pre-processed data of 7 indicators, combined with TOPSIS method, to evaluate the mental health status of minors again, and the result increased by 2.52%.

Therefore, we can conclude that the evaluation model has strong stability.

(2) Model testing of mental health status assessment model

In the process of solving the second problem, we established the evaluation model of minors' mental health status: Based on TOPSIS method, combined with the improved analytic hierarchy process to draw seven indicators weight, calculated each minor mental health score and, in turn, sum, average, the final result is 0.5582, as the region all the minor mental health levels of reflection, thus we analysis, minor mental health status in the region.

We carried out an evaluation test on the above evaluation model: we used the data obtained from the questionnaire survey to take average values of offline grades and total grades, and the results were 3.57 and 28.02 respectively. It can be seen that when the offline score is 28.02 and the total score is 3.57, the corresponding grade is C, indicating that the minors in this area have certain psychological pressure and are vulnerable to external factors, and their overall mental health status is average. The results are consistent with the above model, and the model is verified.

## 6. OUR ADVICE

In the current society, teenagers suffer from less life training, it is difficult to accept pressure from life, and they lack certain coping ability. Their psychological construction mainly comes from three support systems: teachers, parents and peers.

For the school, we think, first of all, we should organize as many as possible for students to organize some mental health classes and group activities, strengthen the communication between students and teachers and classmates, pour out their hearts, relieve mood, relax body and mind; Secondly, appropriate for students to reduce learning pressure, let students have a relaxed and happy learning environment, to avoid class comparison, grade volume, vicious competition, etc.; Finally, I hope schools can open more classes for parents to teach correct family education views, so as to reduce the distance between parents and students, pay more attention to students' mental health, and fundamentally laugh out students' psychological disorders.

For parents, we believe, the best thing to do is to understand. It is inevitable that minors and adults have different perspectives and ways of looking at things. We hope that parents can sincerely listen to what their children want to say, tolerate their children's mistakes and troubles, manage their emotions as much as possible, reduce

contrast, increase encouragement, respect their children's ideas, and consider problems from their perspective. The above are our suggestions on the protection of minors' mental health. We believe that as long as we work together, minors will have a better living and learning environment!

#### REFERENCE

- [1] Ruding. Research on student achievement data analysis and prediction modeling based on K-means clustering algorithm [J]. Microcomputer application, 2021, 37 (05): 148-150.
- [2] Wang Jin, Yan Deju. Using improved k-means clustering algorithm to analyze and evaluate students' performance [J]. Journal of Mudanjiang Normal University (NATURAL SCIENCE EDITION), 2021 (02): 20-22.
- [3] Xie Lin, Li Hongwei, Yuan Yue, Zhou Hailin. Station line loss calculation method based on K-means clustering and improved multi classification correlation vector machine [J]. Journal of electrical engineering, 2021, 16 (01): 62-69.
- [4] Zhang Haiyu, Zhang Guofeng, Li Ming, Luo Zhongyou, Zhao Libin. Data evaluation model of wind solar energy storage power station based on improved analytic hierarchy process [J]. Electronic design engineering, 2021, 29 (10): 32-36.
- [5] Miao Haijun. Research on optimization model of coal mining method based on Improved Analytic Hierarchy Process TOPSIS [J]. Tongmei technology, 2021 (02): 15-17.
- [6] He Xuan, Gao Chong, Cao Huazhen, Li Yang, Yu Tao. Distribution network index evaluation based on improved analytic hierarchy process [J / OL]. Electrical measurement and instrumentation: 1-8 [2021-06-05] <http://kns.cnki.net/kcms/detail/23.1202.TH.20201126.1801.017.html>.
- [7] Li Shaopeng, Zhao Heng, Wang Fuqiang, Yang Dongming. Evaluation of water resources carrying capacity in Jiangsu Province Based on AHP TOPSIS Model [J]. Water resources protection, 2021, 37 (03): 20-25.
- [8] Chen Lei, Wang Zhongli, Xue Yanning, Cui Xiao, Zhao Hui. Power customer credit evaluation method based on Improved AHP method, CV method and TOPSIS method [J]. Hebei electric power technology, 2021, 40 (02): 29-34.
- [9] Liu Fei, Gong ting. Comprehensive evaluation of high quality development in Hubei Province Based on entropy weight TOPSIS Model [J / OL]. Statistics and decision making, 2021 (11): 85-88 [2021-06-05] <https://doi.org/10.13546/j.cnki.tjyjc.2021.11.018>.
- [10] Mao Huimin. Financial risk assessment of real estate enterprises based on entropy weight TOPSIS Model [J]. China market, 2021 (16): 156-159 + 171.
- [11] Yao Jing, Wang Tongshan, Chen Zhe, Yu Zhenzhen. Cruise comfort evaluation based on entropy weight TOPSIS method [J]. China water transportation, 2021 (05): 36-39.

# Value Analysis of Users' Consumption Behavior Based on Stepwise Logistic Regression Model

Yiyao Liu<sup>1, 2\*</sup>, Sijia Zhang<sup>1, 2</sup>, Qian Liu<sup>1, 2</sup>

<sup>1</sup>Engineering Computing and Simulation Innovation Lab, North China University of Science and Technology, Tangshan 063210, China;

<sup>2</sup>School of Science, North China University of Science and Technology, Tangshan 063210, China

\*Corresponding Author.

**Abstract:** In the era of Internet big data, major companies have been able to widen customer acquisition channels by discerning the value of users and formulating specialized marketing strategies for them to achieve small cost marketing to improve user conversion rates. In this paper, based on the behavioral information data of a company's users, a stepwise logistic regression model is used to predict the consumption behavior of users. First, we visualize the distribution of users' cities and logins, and analyze that the users of this enterprise are mainly concentrated in new first-tier cities and mainly new users. After that, a stepwise logistic regression model was used to eliminate insignificant user information and build a prediction model of user consumption behavior to effectively determine the value of users. Finally, using the prediction model of user consumption behavior and the behavior data of already consumed users, the value analysis of user consumption behavior was conducted. The difference percentage pile-up chart of each information between purchased users and all users is used to visualize the factors affecting users' decisions, and a point system incentive system is proposed to enterprises to achieve the goal of improving user conversion rate.

**Keywords:** Stepwise logistic regression model; User value judgment; Visualization presentation; Point Incentive System

## 1. THE BACKGROUND OF THE PROBLEM

Today, with the continuous development of the Internet, companies in all fields are expanding their Internet customer acquisition channels to bring in fresh and active users for their products, increase their desire to buy products, and enhance their brand influence. However, how to identify high-quality users and channels and optimize marketing costs has always been a pain point for companies[1]. This requires analysis of user behavior data to identify the value of users, and then develop special marketing strategies for users to achieve small cost promotions and increase user conversion rates.

## 2. VISUAL PRESENTATION OF USER INFORMATION

### 2.1 USER CITY DISTRIBUTION

According to the "2019 City Business Attractiveness Ranking" published by the First Financial New Tier Cities Research Institute, cities are divided into five categories: "first-tier cities, new first-tier cities, second-tier cities, third-tier cities, and cities below the third tier", and the existing user The cities of users in the existing user

information data are aggregated and classified to obtain the city distribution of users in Figure 1. We can see the weight of the users in each city of the company. From Figure 1, it can be seen that the users of the enterprise are mainly distributed in the new first-tier cities and the cities below the third tier, and the users in the second-tier cities and the third tier cities are equal. Among them, the distribution of users in Tier 1 cities is the least, and the user stickiness of this enterprise in Tier 1 cities is low, and the user stickiness in new Tier 1 cities is higher.

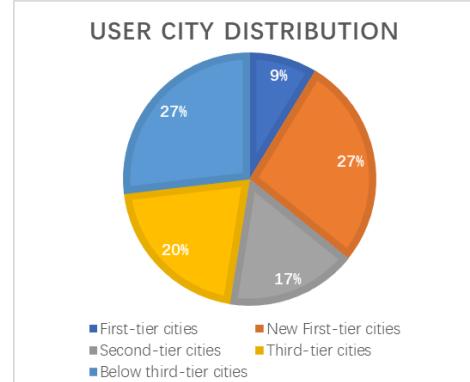


Figure 1.User city distribution

In addition, the bar chart and heat map presented by the 3D map can visually show the distribution of users in each city, and it can be seen from Figure 2 that the users of the enterprise are distributed in all provinces, but the city with the most users is Chongqing and the users are mainly distributed in the central and eastern regions and coastal cities in China. It shows that the users of the enterprise have strong regionalism.

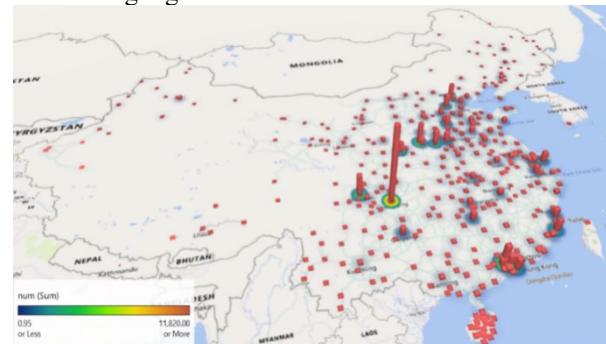


Figure 2.Three-dimensional maps

### 2.2 USER LOGIN STATUS

From Figure 3, we can see that the most users logged in for 5 consecutive days, accounting for 15.85%, and the amount of users logged in for 3 and 6 consecutive days is comparable, accounting for 14.50% and 14.98%

respectively, with very few users logged in for more than 8 consecutive days. It can be analyzed that the users of this enterprise have a certain degree of activity, but the amount of loyal users is small.

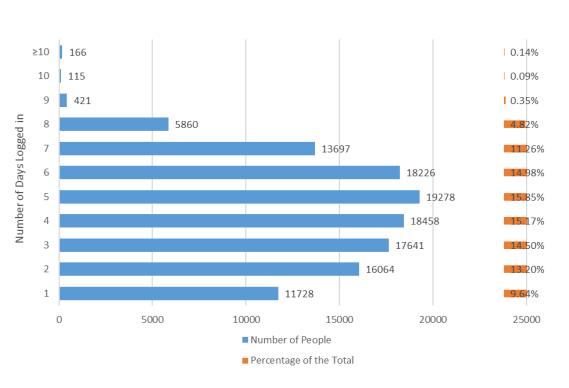


Figure 3. User Login Status

### 3. BUILD A STEPWISE LOGISTIC REGRESSION MODEL

#### 3.1 BRIEF DESCRIPTION OF LOGISTIC REGRESSION MODEL

Logistic regression is a generalized linear regression, which is often used in dichotomous problems and contains two types of variables: dependent variable and independent variable, where the dependent variable is a binary categorical variable, and the dependent variable in this paper is whether the user buys or not, and the independent variable presents the information and behavior of the user[2].

Let the logistic regression have  $r$  independent variables, denoted by  $x_1, x_2, \dots, x_r$ , and the dependent variable  $y \in \{1, 0\}$  denotes the event that the user buys or does not buy,  $y = 1$  means the user buys and  $y = 0$  means the user does not buy.  $y = 1$  practice is expressed in terms of probability  $p$ , and its formula is:

$$p = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_r x_r)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_r x_r)} \quad (1)$$

Where,  $\beta_0, \beta_1, \dots, \beta_r$  are the model parameters,  $\beta_0$  is the intercept term, and the simplified formula by conversion is:

$$p = \frac{1}{1 + e^{-z}} \quad (2)$$

Where,  $z = \beta_0 + \beta_1 x_1 + \dots + \beta_r x_r$ .

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \dots + \beta_r x_r$$

Where,  $1 - p$  is the probability that the user does not buy,  $p$  is the ratio of the likelihood of a user not purchasing an event to the likelihood of purchasing an event, known as *odds*. The linear function is obtained by taking the logarithm of *odds*, and logistic regression is used to optimize the model by finding the best parameters  $\beta_0, \beta_1, \dots, \beta_r$ .

The maximum likelihood estimation function is used to measure the  $\beta_0, \beta_1, \dots, \beta_r$  model parameters.

model parameters, with  $m$  sets of observations, the maximum likelihood function is:

$$L(\beta) = \prod_i^m p(y^{(i)} | x^{(i)}; \beta) = \prod_i^m h_\beta(x^{(i)})^{y^{(i)}} [1 - h_\beta(x^{(i)})]^{1-y^{(i)}} \quad (3)$$

Where,  $h_\beta(x^{(i)}) = \frac{1}{1 + e^{-\beta^T x^{(i)}}}$ , model predictions for the  $i$ -th sample;  $\beta_0, \beta_1, \dots, \beta_r$ , The set of parameters consisting of  $\beta_0, \beta_1, \dots, \beta_r$ , to facilitate the calculation of the derivative, the log-likelihood function is established:

$$\ln L(\beta) = \sum_i^m y^{(i)} \ln[h_\beta(x^{(i)}; \beta)] + \{1 + y^{(i)} \ln[1 - h_\beta(x^{(i)}; \beta)]\} \quad (4)$$

where, the estimates of the logistic regression coefficients that maximize the log-likelihood function are obtained by finding the partial derivatives of the parameters  $k$ , respectively.

#### 3.2 STEPWISE LOGISTIC REGRESSION MODEL

Wald is the value of the statistic for the regression coefficient test:

$$Wald = \left( \frac{B}{S.E.} \right)^2 = \left( \frac{B}{\sqrt{D(\beta_j)}} \right)^2 \quad (5)$$

Sig. is the probability of significance of the Wald test.

In the logistic regression model, if the coefficient of an independent variable does not pass the significance test, i.e., the p-value corresponding to that independent variable is greater than 0.1, the independent variable is considered to have insignificant influence on the model building process, and the variable will not be included in the final model. Based on this, the stepwise logistic regression model can generate fewer explanatory variables[3].

In this paper, a backward stepwise logistic regression model is chosen to model all regression independent variables and perform the Wald coefficient test, and if it passes the test, the independent variable is introduced into the regression model, which is recorded as the set of selected variable indicators. If the significance of the independent variable does not pass the Wald test, the independent variable is excluded. With the elimination of the less significant independent variables, smaller AIC values can be obtained and the model is simplified, and the error of the simplified model on the validation set is smaller than that of the full-variable model, and the stepwise logistic regression model is finally obtained.

Variable selection was performed using stepwise logistic regression to generate a simplified model with fewer explanatory variables according to the AIC criterion, and a new model with 29 explanatory variables was obtained. Twenty regression coefficients with p-values that did not contribute significantly to the equation were excluded and the model was refitted, and each coefficient of the new model was highly significant ( $p < 0.05$ ), and the independent variables of the fitted model are shown in Table 1.

Variables in the equation					
	B	Standard Error	Wald	Degrees of freedom	Significance
Step 1	login_day	-.292	.022	169.323	1 .000
	login_diff_time	.070	.009	64.309	1 .000
	distance_day	-.015	.001	847.232	1 .000
	login_time	.003	.000	49.130	1 .000
	chinese_subscribe_num	1.065	.053	400.688	1 .000

math_subscribe_nu_m	.330	.075	19.327	1	.000
add_friend	5.285	1.754	9.077	1	.003
camp_num	-.087	.032	7.300	1	.007
learn_num	-.353	.032	122.775	1	.000
finish_num	.158	.032	24.730	1	.000
study_num	1.318	.063	443.254	1	.000
coupon	1.382	.031	1935.02	1	.000
course_order_num	.761	.032	572.284	1	.000
first_order_time	.000	.000	58.317	1	.000
first_order_price	.021	.004	30.384	1	.000
age_month	.005	.002	10.046	1	.002
main_home	.005	.002	8.389	1	.004
main_home2	.003	.001	8.342	1	.004
mainpage	.035	.005	42.208	1	.000
schoolreportpage	.021	.007	8.633	1	.003
lightcoursetab	-.060	.016	14.891	1	.000
evaluationcenter	-.024	.007	11.206	1	.001
coupon_visit	-.538	.015	1274.47	1	.000
task	.025	.005	24.785	1	.000
video_read	.009	.002	19.003	1	.000
answer_task	-.024	.003	64.969	1	.000
chapter_module	-.015	.003	23.710	1	.000
share	.016	.006	8.448	1	.004
constants	892.46	14861.80	.004	1	.952
	6	0			

Table 1. Variables in the equation

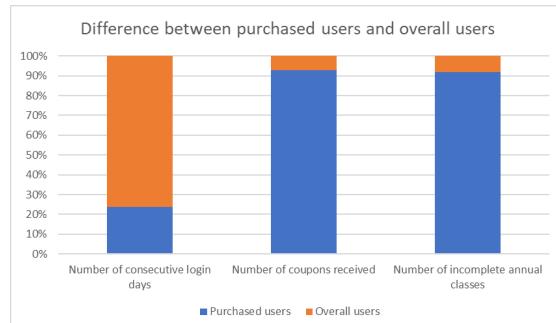
The stepwise logistic regression model built using some of the independent variables was 99.6% correct in predicting user non-purchase events and 47.6% correct in predicting user purchase events, with a model accuracy of 97.9%. The correct rate of 47.6% can be agreed because the data of purchasing users in the user behavior data set is too small compared with the data of non-purchasing users. The prediction classification table of the model is shown in Figure 2.

Classification table					
Real measurement	Predicted				
	result		Percent correct		
	0	1			
Step 1 result	0	86889	332	99.6	
	1	1583	1438	47.6	
Overall percentage			97.9		

Table 2. Classification table

#### 4. SOUND ADVICE FOR COMPANY

Through the established stepwise logistic regression model, the variables with high significance were screened and presented visually to plot the percent difference pile-up of purchased users versus all users, as shown in Figure 3.



From Figure 3, we can see that the purchased users generally have fewer consecutive login days, more coupons and more users who have not completed their annual classes[4]. Therefore, enterprises can appropriately increase the number of coupons issued and set up point incentive rules at the same time. For users who have logged in for more consecutive days and completed the annual course, coupons can be issued to increase users' stickiness to the course.

#### REFERENCES

- [1] Han Kang Kang. Research on the prediction of microblog users' retweeting behavior based on retweeting habits[D]. Hebei University, 2020.
- [2] Cao J, Zhang Yansong, Liu Q, Yang WJ, Gao Feng, Liu Zengxia. Research on supplier risk model of oil companies based on logistic regression score card[J]. Oil and Gas and New Energy, 2021, 33(05):51-57.
- [3] Jiaxin Li. Analysis of personal credit evaluation based on classification algorithm under stepwise logistic regression[J]. Journal of Hunan Academy of Arts and Sciences (Natural Science Edition), 2021, 33(01):5-8+57.
- [4] Hou Baozhu. On the study of establishing a point system incentive mechanism[J]. Economic and Trade Practice, 2018(20):230+232.

# NCFA: Networks Correlations Forecast and Assessment of Music Evolution

Yibo Wang<sup>1, 2</sup>, Yunjiang Han<sup>3</sup>, Yi Zhu<sup>1, 2</sup>

<sup>1</sup>Engineering Calculation and Simulation Innovation Lab, North China University of Science and Technology, Tangshan 063210, China;

<sup>2</sup>Yisheng Innovation Education Base, North China University of Science and Technology, Tangshan 063210, China;

<sup>3</sup>College of Mechanical and Electrical Engineering, Shenyang Aerospace University, Shenyang 110135, China

\* Corresponding Author:

**Abstract:** Musicians' styles, genres and musical characteristics may be influenced by the times, environment and other musicians. Firstly, Spelman correlation coefficient was used to analyze the correlation between different genre to determine the degree of similarity between genre and within genre. At the same time, the representative musicians who choose to distribute more songs calculate correlation coefficient to analyze the correlation between them. In the study of different types over time, the factor analysis model is used to analyze 14 evaluation indexes into two-dimensional evaluation index models, and then analyzed 2 new evaluation indexes corresponding to 20 genres over time. Finally, the passion vitality index of each genre is rising and the elegant positive index is decreasing. Moreover, with the help of ARIMA model, We explored the changes of music characteristics of Pop/Rock genre in the evolution of music, established a ARIMA model, and better described the changes of four evaluation indexes of Pop/Rock genre from 1928 to 2020. Finally, pop/rock school's energy index, dance, popularity index, growing, enthusiasm has declined conclusion. We finally conduct

sensitivity analysis, dissect pros and cons of our model and present a memo of our work to the ICM Association.

**Keywords:** ARIMA model; Correlation Analysis; Factor analysis; Networkx

## 1 INTRODUCTION

Music is a part of human society and an important part of cultural heritage, while musicians are an important carrier to create music and integrate it into human society. The style, genre and musical characteristics of a musician may be influenced by the times, environment and other musicians. These factors which may affect the style, genre and musical characteristics of musicians are the important basis for judging the development and evolution of musical genres. By reasonably quantifying these data, we can better understand how music evolves in society over time.

In this work, we propose a novel framework, named NCFA (Networks, Correlations, Forecast and Assessment of Music Evolution), to fill in the gap above. The framework of NCFA is shown in Figure 1.

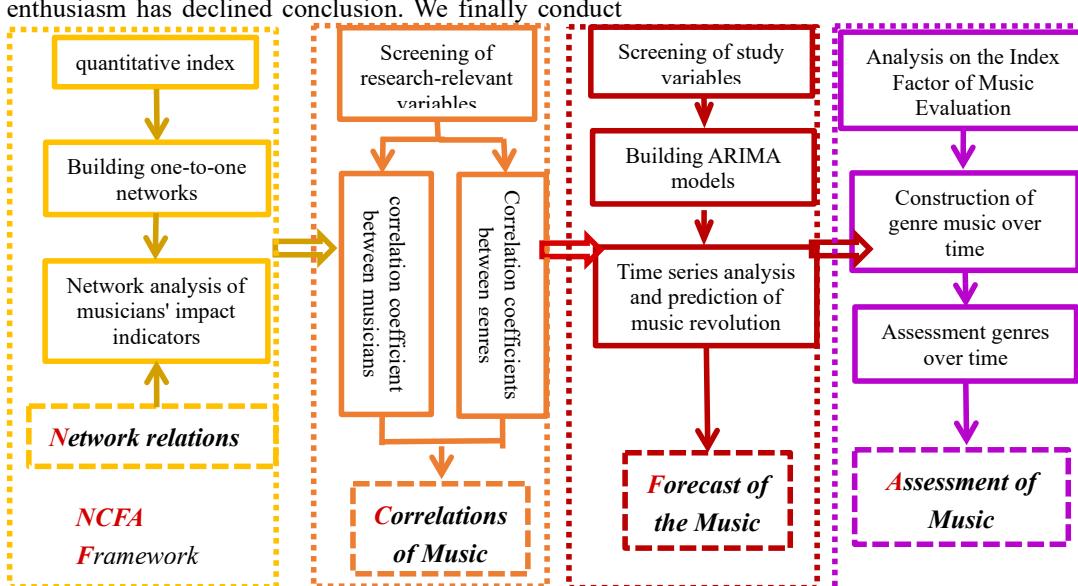


Figure 1: Framework of CAFE

## 2 MODELING AND PROBLEM HANDLING

A total of 20 genres, 9 genres of musicians and active time were provided in influence\_data. The data given from the data set can show that the genre choice of the follower is influenced by the music genre of the influence to a certain

extent. In this regard, we screened and analyzed the genres of influence of 20 genre musicians given by the data set, except for the genres whose results were not universal because of the small number of genre musicians. Finally, 5 musicians were retained in more than 1, 000 genres for

statistical analysis.

The change trend of the proportion of musicians of different genres at any time is as follows:

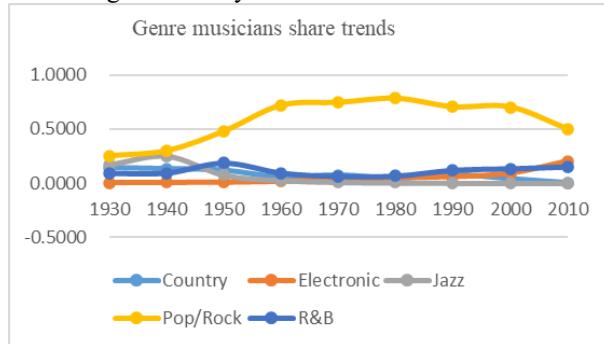


Figure 2 Genre musicians share trends

In addition to the times can be an important factor affecting musicians' genres, the music genres in which they are influenced will also have a profound impact on musicians' own genres. In this regard, we made a statistical analysis of the selected five typical musical genres and their influencers, and obtained the following quantitative results:

We can see that most of the followers are influenced by their influence genres. There is an anomaly in the followers of Electronic genres, and it is preliminarily concluded that the reason is that their genres Electronic out of Pop/Rock, so the development of their own genres is deeply influenced by the Pop/Rock genres, and in the later Vacal genres, This phenomenon also shows obvious similarities.

Because the original data set contains 13 music evaluation indicators, and there is a correlation between these music evaluation indicators.<sup>[3]</sup>Therefore, the 13 music evaluation indicators are analyzed by factor analysis, that is, dimension reduction.

The existing n samples, p indicators, can form a sample matrix of n \* p size:

$$x = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} = (x_1, x_2, \dots, x_p)$$

$x_1, x_2, \dots, x_p \Rightarrow f_1, f_2, \dots, f_p$  ( $m \leq p$ ) Among them, and they meet:

$$\begin{cases} x_1 = u_1 + a_{11}f_1 + a_{12}f_2 + \dots + a_{1m}f_m + \varepsilon_1 \\ x_{21} = u_2 + a_{21}f_1 + a_{22}f_2 + \dots + a_{2m}f_m + \varepsilon_2 \\ \vdots \\ x_p = u_p + a_{p1}f_1 + a_{p2}f_2 + \dots + a_{pm}f_m + \varepsilon_p \end{cases}$$

$x = u + Af + \varepsilon$  As a result, it is assumed that:<sup>[4]</sup>

$$\begin{cases} E(f) = 0 \\ E(\varepsilon) = 0 \\ Var(f) = I \\ Var(\varepsilon) = D = diag(\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2) \\ cov(f, \varepsilon) = E(f\varepsilon') = 0 \end{cases}$$

By using the spss25 software to analyze the factors, the gravel map corresponding to the eigenvalues of each factor is first drawn, as shown in figure 3.

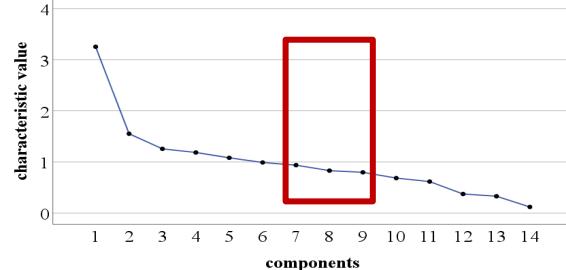


Figure 3 Gravel map

It can be seen from figure 6 that the curve between component 1 and component 2 is steep, while the curve after component 2 is relatively smooth, so the retention component 1 and component 2 are analyzed.

After determining the retention of two components, multiply the original component matrix with the rotation matrix k,

$$k = \begin{bmatrix} 0.945 & 0.328 \\ 0.328 & -0.945 \end{bmatrix}$$

The composition matrix after rotation is obtained, as shown in table 5:

Table 1 Composition matrix after rotation

	component	
	1	2
danceability	.156	.769
energy	.888	.107
valence	.238	.776
tempo	.330	-.033
loudness	.828	.183
mode	-.086	.130
key	.053	-.010
acousticness	-.838	-.056
instrumentalness	-.345	-.365
liveness	.192	-.287
speechiness	.184	-.067
explicit	.317	-.154
duration ms	.143	-.480
popularity	.619	.016

table 5 shows that, energy, of component 1 High loudness and popularity, Hence the interpretation of component 1 as Passionate Power; danceability, corresponding to simultaneous component two High valence, As a result, component 2 is interpreted as Graceful and positive.

Each factor weight is obtained by factor analysis, the original component matrix X is multiplied by the weight matrix, and the corresponding values of the new evaluation indexes of each genre are obtained:

$$\text{New } X = \begin{bmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \\ \vdots & \vdots \\ \alpha_{14} & \beta_{14} \end{bmatrix} * \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{114} \\ x_{21} & x_{22} & \cdots & x_{214} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{n14} \end{bmatrix}$$

The value of the new evaluation index corresponding to each genre is obtained, and the change of the value of the two new components of each genre with time is analyzed respectively. The following six new evaluation indexes are listed as follows: Country, Folk, Jazz, Latin, Pop/Rock and R&B, respectively.

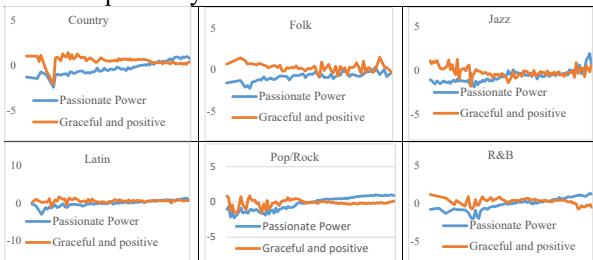


Figure 4 Changes of genres over time

With the passage of time, the Passionate Power of each genre has increased obviously, while the Graceful and positive of each genre has declined over time, especially between 1960s and 1970s

It is concluded that with the passage of time, the energy index, popularity index and passion power index of each genre have all increased, and the dance and enthusiasm of each genre have decreased with time. According to the relevant literature[5], the corresponding events in the 1960s and 1970s include the assassination of Kennedy, the failure of the Vietnam War, the Cuban missile crisis, and the assassination of Martin Luther King. Therefore, We determined that the negative enhancement of music in this era may be related to these events, and in order to find comfort from these negative events, musicians began to create music with passionate power to alleviate their negative emotions. Therefore, the passion of this era, strength indicators are still increasing.

### 3 STUDIES OF GENRES OVER TIME

#### 3.1 Selection of research objects and variables

To analyze the evolution of music genres over time, Pop/Rock genres are selected to study because of the large proportion of Pop/Rock genres in the full\_music\_data data set. Among the 14 music evaluation indexes, some music evaluation indexes such as explicit, duration\_ms are fixed at each time point, or can not describe the music changes, so there is no research significance. Based on this method of selecting music evaluation index, this paper studies 4 of 14 music evaluation indexes, namely: danceability, energy, valence, popularity.

#### 3.2 Establishment of difference autoregressive moving average model

The four variables are classified and summarized in the full\_music\_data data set, and the average values of these four indexes in each year of the Pop/Rock genres from 1928 to 2020 are obtained.

$x$  is the time series data to be studied  $x_1, x_2, \dots, x_t$ . The ACADEMIC PUBLISHING HOUSE

covariance is assumed to be stationary, that is, to satisfy the stationarity of the time series. Based on this assumption, We uses the expert modeler in the SPSS 25 to model the time series. The expert modeler automatically finds the best fit model for each dependent sequence. If an independent variable is specified, the expert modeler selects those models that have a statistically significant relationship with the dependent sequence for the content in the ARIMA model. When appropriate, use difference, square root or natural logarithmic conversion to transform model variables. By default, the expert modeler considers both exponential smoothing and ARIMA models.<sup>[7]</sup> The four time series model types corresponding to the four evaluation indexes established by the expert modeler are shown in the following table:

Table 2 Type of model

Model ID	Types of models			
	Danceability	Model 1	ARIMA(0, 0, 2)	
	Energy	Model 2	ARIMA(2, 1, 0)	
	Valence	Model 3	ARIMA(0, 1, 1)	
	Popularity	Model 4	ARIMA(1, 1, 0)	

ARIMA model parameters given are shown in the following table:

Table 3 ARIMA model parameter

	estimate	standard error	T	conspicuousness
Danceability quantity	.375	.010	35.912	.000
	MA delay 1	-.420	.091	-4.620 .000
	delay 2	-.315	.092	-3.415 .001
	AR delay 1	-.579	.130	-4.466 .000
Energy quantity	difference	.334	.120	-2.774 .007
	difference	1		
	MA delay 1	.756	.036	21.250 .000
Valence quantity	difference	1		
	AR delay 1	-.428	.084	-5.093 .000
	difference	1		
Popularity quantity	difference	1		
	MA delay 1	.656	.096	6.844 .000
	AR delay 1	-.428	.084	-5.093 .000

First, the original time series model is established:<sup>[8]</sup>

$$(1 - \sum_{i=1}^p \alpha_i L^i)(1 - L)^d y_t = \alpha_0 + (1 + \sum_{i=1}^q \beta_i L^i) \varepsilon_t$$

The L is the y lag operator, which satisfies:  $L^i y_t = y_{t-i}$

Take the Danceability model for example, because Danceability is ARIMA(0, 0, 2) the model, the p, d is 0, so the model expression is:

$$y_t = \alpha_0 + (1 + \beta_1 L + \beta_2 L^2) \varepsilon_t$$

$$\text{Simplified: } y_t^2 - \alpha_0 y_t = \alpha_0 \varepsilon_t \beta_1 y_{t-1} + \alpha_0 \varepsilon_t \beta_2 y_{t-2}$$

Similarly, the time series ARIMA models corresponding to Energy, Valence, Popularity three evaluation indexes are obtained:

$$y_t = \frac{\alpha_0 + \varepsilon_t}{1-d} + \alpha_1 y_{t-1} + \alpha_2 y_{t-2}$$

$$y_t - \beta \varepsilon_t \frac{y_{t-1}}{y_t} = y_{t-1} + \alpha_0 \varepsilon_t$$

$$y_t + \frac{\alpha y_{t-1}^2}{y_t} = (\alpha + 1)y_{t-1} + \alpha_0 + \varepsilon_t$$

After inserting the parameters, we can gain:

$$y_t = 0.375 + (1 + \beta_1 L + \beta_2 L^2) \varepsilon_t$$

$$y_t = \frac{0.006 + \varepsilon_t}{1-d} - 0.579 y_{t-1} - 0.334 y_{t-2}$$

$$y_t - \beta \varepsilon_t \frac{y_{t-1}}{y_t} = y_{t-1} + 0.756 \varepsilon_t$$

$$y_t - \frac{0.428 y_{t-1}^2}{y_t} = (-0.428 + 1)y_{t-1} + 0.656 + \varepsilon_t$$

### 3.3 Analysis of time series models

According to the predicted value, draw outdanceability, energy, valence, popularityEffect diagram of actual value and predicted value corresponding to 4 indexes:

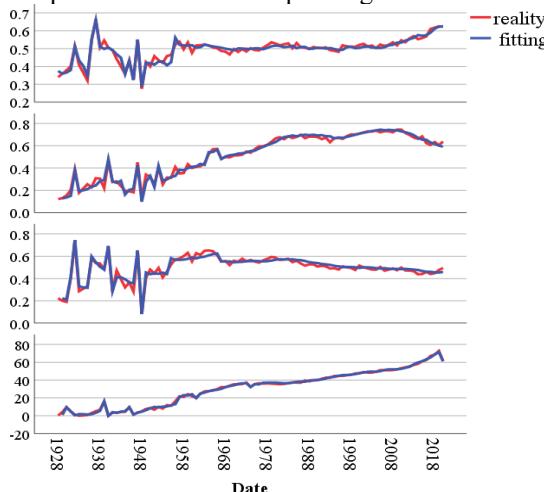


Figure 5 Results of the forecast

It can be found from the diagram that the ARIMA model fits well with the actual value and can reflect the changing trend of the four indexes with time in most Pop/Rock genres.

Synthesizing above analysis, from time series change trend chart, we can see that the dance sex of Pop/Rock song has been increasing steadily since 1960, especially from 2010 to now Pop/Rock the dance sex of music has increased obviously; at the same time, in energy aspect, from 1928 to 2010, the energy index of Pop/Rock has been increasing significantly, and from 2010 to now, the energy index has decreased; On the positive side, the enthusiasm of Pop/Rock grew slowly from 1928 to 1968, decreasing slowly from 1968 to present, and the trend was relatively stable; compared with other indicators, the popularity of Pop/Rock increased significantly from 1928 to 2018. However, from 2018 to now, the popularity of Pop/Rock has a significant downward trend.

### 3.4. Use of networks to identify social and political impacts

From the 4.1 link, We has established a networkx network

based on influencers and followers genres and linked influencers and followers. As can be seen from figures 1 and 9, the evolution of musicians' music genres in the data set, as well as the changes in musical characteristics, took place near 1960. The corresponding events in the 1960s and 1970s include the assassination of Kennedy, the failure of Vietnam War, the Cuban missile crisis, and the assassination of Martin Luther King. Therefore, We determined that the negative enhancement of music in this era may be related to these events, and in order to find comfort from these negative events, musicians began to create music with passionate power to alleviate their negative emotions. Therefore, the passion of this era, strength indicators are still increasing.

The above contents have proved that the network established by We can reflect the trend of social development and the degree of social unrest through the quantitative form of the change of musical characteristic index of musicians in music evolution. Therefore, the network can identify the influence of society and politics on music evolution in this way.

### 3 Time series model tests

To test the accuracy of 4 ARIMA models, the self-correlation ACF and partial autocorrelation coefficients PACF: calculated

$$ACF = r_s = \hat{\rho}_s = \frac{\sum_{t=s+1}^T (x_t - \bar{x})(x_{t-s} - \bar{x})}{\sum_{t=1}^T (x_t - \bar{x})^2}$$

$$PACF = x_t = \phi s_1 x_{t-1} + \phi s_2 x_{t-2} + \dots + \phi s_s x_{t-s} + e_t$$

The results are shown in Figure 6:

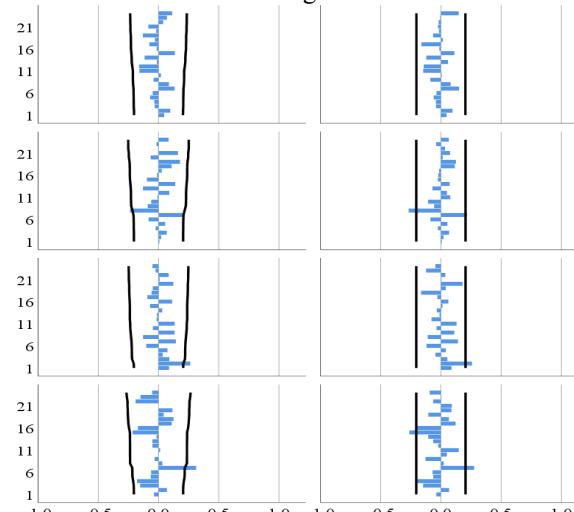


Figure 6 ACF and PACF of residuals

We can see from the ACF of the residuals and the PACF figure that the self-correlation and partial autocorrelation coefficients of most of the lag orders are within the significant level, which indicates that there is no significant difference between them and 0.

In addition, it can be seen from the following table that the values obtained from the residual error detection are all greater than 0.05, which indicates that all the predicted data accept the original hypothesis at 95% confidence level, that is, the residual error is a white noise sequence.

As a result, the ARIMA model can well identify the time series data in this problem.

Table 4ARIMA model testing

Model	fit statistics	Yang -Box Q(18)	Number of outliers
	R smooth	statistics DF	conspicuousness
danceability- Model 1	.922	14.939	16 .529
energy-Model 2	.891	20.325	16 .206
valence-Model 3	.951	21.285	17 .214
popularity- Model 4	.878	29.176	17 .303

#### 4 CONCLUSION

From the network established, we are committed to exploring the influence of follower genres. We conclude that the development of followers is influenced by the corresponding genres and times, and the development of similar genres also has a certain impact on followers. At the same time, we determine the correlation in the study. The musical characteristics of genre musicians are the main factors that affect the evolution of music and lead to the evolution of music. When we further explore the model, we combine the social environment background with the comprehensive development trend of musicians' musical characteristics, which adds variables of social and political factors to our model. The evolution of music reflects the change of social environment, and the change of music characteristics reflects the change of political theme.

#### REFERENCES

[1] Zhang Tinghui. NetworkX - based

implementation of complex network features and its application in biological networks[D]. Institutes Of Technology Of South China, 2016.

- [2] Wang Xiaoyan, Li Americas. On the Correlation Coefficient of Grade and Spearman Grade[J]. Journal of Guangdong Light Industry Vocational and Technical College, 2006(04):26-27.
- [3] Wu Xiaoting, Yan Deloitte. Analysis and Research on Data Dimension Reduction Method[J]. Computer Applied Research, 2009, 26(08):2832-2835.
- [4] Xie Suwen. Application Research of Mathematical Model Based on Principal Component Analysis and Factor Analysis[D]. Shandong University of Technology, 2016.
- [5] Yu Mingchuan. A Study on the Form of Social Music in the Sixties and Seventies of the 20th Century[D]. A Study on the Form of Social Music in the Sixties and Seventies of the 20th Century, 2019.
- [6] Qin Yubing, Deng Chunlin, Yang Liu. Research on the Construction of Network Public Opinion Evaluation Index System Based on Pearson Correlation Coefficient[J]. Intelligence exploration, 2018(10):15-19.
- [7] Cao Tong, Qiao Yu, ARMA algorithm in time series analysis and its software implementation[J]. Journal of Union University of Beijing. 1996, (02).
- [8] First Zhaoyong, Yang Yuanyuan, Modeling Methods and Processes of Time Series Problems[J]. Mathematical Theory and Application. 2012, 32(01).
- [9] Ma Fengming, Wang Zhongli. Analysis and Application of Hypothesis Test Method[J]. Journal of Changchun University, 2012, 22(02):188-192+196.

# Catalytic Experimental Study on Preparation of C4 Olefin from Ethanol

Bingbing Shi\*, Zihang Chen, Yingyi Xie

School of Artificial Intelligence, North China University of Technology, Tangshan 063210, China

\*Corresponding Author.

**Abstract:** Ethanol is a key renewable compound. It is of great value to study the relationship between ethanol conversion and C4 olefin selectivity and temperature, and to design a reasonable combination of catalysts to improve the yield of C4 olefin. In this paper, the effects of different catalyst combinations and reaction temperatures on ethanol conversion and C4 olefin selectivity were studied. The experimental data were preprocessed to obtain relatively complete experimental data. Python software was used to perform linear and nonlinear regression fitting for the relationship between ethanol conversion, C4 olefin selectivity and temperature.

**Keywords:** Olefin Preparation From Ethanol; Regression Model; Fitting Curve

## 1. BACKGROUND

Ethanol is a key renewable compound. It is a hot research topic to synthesize other chemical products with ethanol as raw material. C4 olefin is one of the important petrochemical basic raw materials, and ethanol is the raw material for the production and preparation of C4 olefin. At present, there are few specific studies on improving the selectivity of the gas product propylene in this reaction. It is of great theoretical significance and application value to explore the technological conditions for the preparation of C4 olefin by ethanol catalytic coupling. According to a chemical laboratory for different catalysts at different temperatures to do a series of experimental data.

## 2. PROBLEM ANALYSIS

We preprocessed the experimental data and obtained relatively complete experimental data. Python software was used to perform linear and nonlinear regression fitting on the relationship between ethanol conversion rate, C4 olefins selectivity and temperature, and comparative analysis of the relationship between ethanol conversion rate, C4 olefins selectivity and temperature. Establish a regression fitting model for data changing with time to explore the relationship between variables and time.

## 3. ASSUMPTIONS OF THE MODEL

In order to establish a simplified model, the team simplified each participant of the model as follows during the modeling process:

1. In the process of preparing C4 olefin by ethanol coupling, all the experimental processes are legal and standardized, and no safety accidents will occur.
2. It is assumed that the density of materials does not change during the preparation of C4 olefin by ethanol coupling.
3. It is assumed that there are no other side reactions during the preparation of C4 olefin by ethanol coupling,

and the utilization rate of carbon atom is 100%.

4. It is assumed that the reaction vessel used in the preparation of C4 olefins by ethanol coupling is of the same size and quality.

5. The quality changes before and after the reaction of preparing C4 olefin by ethanol coupling are not considered.

## 4. DEFINITION AND NOTATION

In order to simplify problem analysis and data processing, symbols are stipulated as follows:

symbol	define
$\alpha$	Ethanol conversion
$\beta$	C4 olefins selectivity
$T$	Reaction temperature
$t$	The reaction time
$s_i$	Catalyst combination variable

## 5. THE ESTABLISHMENT AND SOLUTION OF MODEL

### 5.1 DATA PREPROCESSING

The data table shows the experimental information of catalysts numbered A1 ~ A14 and B1 ~ B7. First of all, through our preliminary study, the catalyst combination of GROUP A9 and group B5 and group A12 and group B1 is the same. However, due to the different charging methods of i and ii, there are only a few experimental groups, so it is temporarily impossible to determine the influence of charging methods on products. Consider them as different catalyst groups. Secondly, for the data in these information, data cleaning is essential, and it is necessary to eliminate useless information and outliers.

Step 1: remove the abnormal data, according to annex 1, finishing all 21 different combination of catalyst solution found that A11 catalyst composition compared with other groups, catalyst carrier for quartz sand, without the presence of HAP, ethanol conversion rate and selectivity of C4 olefin is lower, the problem of the whole inquiry no larger value, thus eliminating.

Step 2: Complete the missing data -- after screening out the worthless data, we found that in the experimental data, the temperature setting interval was different under different catalyst combinations. In order to simplify the problem, we set 25°C as the setting interval to study the data within the range of 250 ~ 400°C. For intermediate data missing, we use the before and after average method to complete the median; For data missing before and after, we use the third order exponential smoothing method to complete the value.

Before and after average method: the numerical average of the former and the latter of the missing data is used as the compensating data of the missing value.

Third-order exponential smoothing method: Given the

smoothing coefficient  $U = 0.4$ , the calculation formula of cubic exponential smoothing is as follows:

$$\begin{cases} S_t^{(1)} = ux_t + (1-u)S_{t-1}^{(1)} \\ S_t^{(2)} = uS_t^{(1)} + (1-u)S_{t-1}^{(2)} \\ S_t^{(3)} = uS_t^{(2)} + (1-u)S_{t-1}^{(3)} \end{cases} \quad (1)$$

Predict the value of future TD period. The calculation formula of  $x_{t+TD}$  is:

$$x_{t+TD} = A_{TD} + B_{TD}TD + C_{TD}TD^2 \quad (2)$$

Among them:

$$A_t = 3S_t^{(1)} - 3S_t^{(2)} + S_t^{(3)}$$

$$B_t = \left(\frac{u}{2(1-u)^2}\right)[(6-5u)S_t^{(1)} - 2(5-4u)S_t^{(2)} + (4-3u)S_t^{(3)}] \quad (3)$$

$$C_t = \left(\frac{u^2}{2(1-u)^2}\right)[S_t^{(1)} - 2S_t^{(2)} + S_t^{(3)}]$$

## 5.2 MODEL PREPARATION

The relationship between Y and X is a correlation, that is, when the independent variable X changes, the dependent variable Y changes roughly in accordance with a certain law. The relationship between the two cannot be seen intuitively and needs to be determined by statistics. Regression analysis is a mathematical statistical method to study the relationship between variables in random phenomena.

Let's use the least square method to solve the regression equation

$$b = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (4)$$

After the linear regression of the model, the goodness of fit of coefficients of regression model is evaluated by determination coefficient to check whether the regression equation has practical significance.

$$R^2 = \frac{\text{Return to the sum of squares}}{\text{The total sum of squares}} = 1 - \frac{\text{Sum of squares of residuals}}{\text{The total sum of squares}} \quad (5)$$

The closer  $R^2$  is to 1, the better the regression fit of x and y is.

## 5.3 ANALYZE THE RELATIONSHIP BETWEEN

Table 1. Data table of fitting graph in A method

Catalyst combination number	A1	A2	A3	A4	A5	A6	A7
b	0.32		0.48	0.58	0.41	0.5	0.38
R <sup>2</sup>	0.966	0.958	0.973	0.995	0.873	0.966	0.999
Catalyst combination number	A8	A9	A10	A11	A12	A13	A14
b	0.34	0.25	0.18	0.2	0.28	0.25	0.33
R <sup>2</sup>	0.947	0.82	0.823	0.782	0.913	0.853	0.916

The fitting diagram of ethanol conversion rate and temperature for each catalyst in charging mode II (B1 ~ B7) is as follows (B2 ~ B4), and see appendix for details.

Table 2. Data table of fitting graph under B methods

Catalyst combination number	B1	B2	B3	B4	B5	B6	B7
b	0.28	0.27	0.13	0.21	0.27	0.38	0.42

As can be seen from the fitting diagram of the relationship between ethanol conversion rate and temperature, for all catalyst combinations given in Annex I, the change rule of ethanol conversion rate is always increasing with the increase of temperature, and the fitting degree is good according to the determination coefficient R<sup>2</sup>.

Table 3. Data table of fitting graph in different ways A

ACADEMIC PUBLISHING HOUSE

## VARIABLES BASED ON REGRESSION FITTING

### 5.3.1 RELATIONSHIP BETWEEN ETHANOL CONVERSION RATE AND TEMPERATURE

The fitting diagram of ethanol conversion rate and temperature for each catalyst in charging mode I (A1~A14) is as follows (A1~A3)

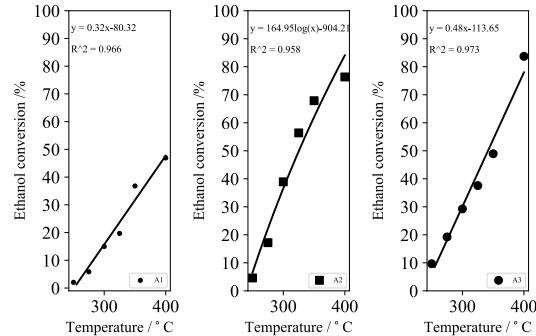


Figure 1. Fitting diagram of the relationship between ethanol conversion rate of A1~A3 and temperature Among them, A2 catalyst combination simulates the nonlinear regression equation.

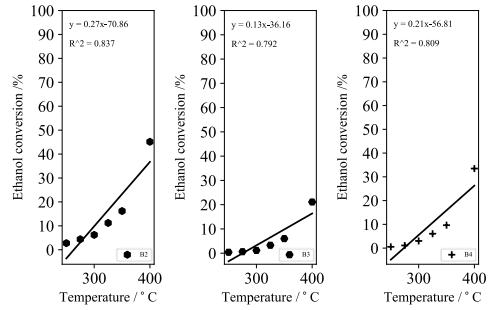


Figure 2. Fitting diagram of b2-B4 ethanol conversion rate and temperature

The slope and determination coefficient of the fitting straight line diagram of each catalyst in charging mode I (A1~A14) are as follows:

### 5.3.2 RELATIONSHIP BETWEEN C4 OLEFINS SELECTIVITY AND TEMPERATURE

For each catalyst in charging mode I (A1~A14), the relationship between the selectivity of C4 alkenes and temperature is illustrated as follows (A1~A3).

Catalyst combination number	A1	A2	A3	A4	A5	A6	A7
b		0.23	0.34	0.23	0.23	0.2	0.19
R <sup>2</sup>	0.59	0.926	0.986	0.917	0.94	0.747	0.926
Catalyst combination number	A8	A9	A10	A11	A12	A13	A14
b	0.24	0.25	0.05	0.05	0.2		0.14
R <sup>2</sup>	0.981	0.995	0.703	0.974	0.961	0.98	0.904

Among them, A1 and A13 catalysts were combined to simulate the nonlinear regression equation.

The fitting diagram of ethanol conversion rate and Table 4. Data table of fitting graph under B methods

Catalyst combination number	B1	B2	B3	B4	B5	B6	B7
b	0.24	0.24	0.12	0.1	0.15	0.19	0.23
R <sup>2</sup>	0.968	0.965	0.943	0.801	0.956	0.965	0.989

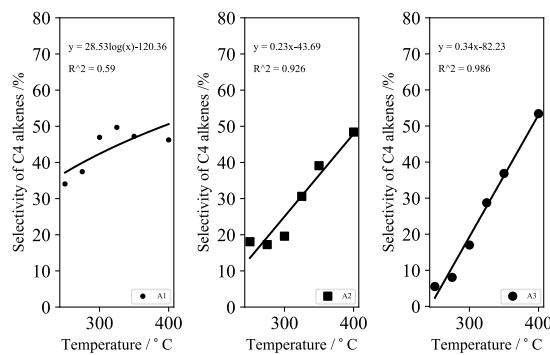


Figure 3. fitting diagram of A1~A3 C4 olefins selectivity and temperature

As can be seen from the fitting diagram of the relationship between the selectivity of C4 olefin and temperature, for all catalyst combinations given in Annex I, the selectivity of C4 olefin keeps increasing with the increase of temperature, which can be seen from the determination coefficient R<sup>2</sup> that the fitting degree is good.

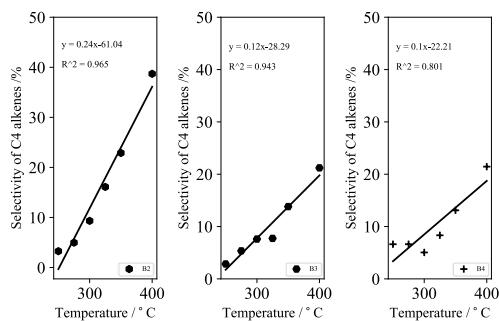


Figure 4. B2 ~ B4 C4 olefin selectivity and the relationship between temperature fitting chart

### 5.3.3 ANALYSIS OF RELATIONSHIP BETWEEN REACTION PRODUCTS AND TIME

The regression fitting analysis of the test results of a given catalyst combination at 350 degrees at different times in an experiment is as follows:

As can be seen from the figure, when the temperature is 350 degrees, the conversion rate of ethanol decreases slowly with the increase of time under the catalysis of a given combination of catalysts. The selectivity of ethylene, C4 olefin, acetaldehyde, fatty alcohols with carbon number 4-12, methylbenzyl alcohol and other products

temperature for each catalyst in charging mode II (B1 ~ B7) is shown in appendix, with examples (B2 ~ B3).

gradually reached a state of dynamic equilibrium over time.

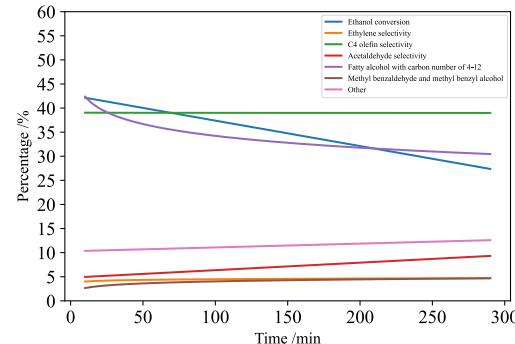
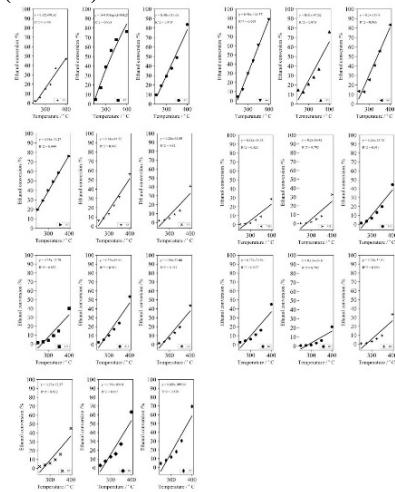


Figure 5. Fitting diagram of test results of a given catalyst combination at 350 degrees at different times in an experiment

### 6. THE APPENDIX

The fitting figure of ethanol conversion rate and temperature for each catalyst (A1~B7) is shown below:

The fitting figure of the relationship between the selectivity of C4 alkenes and temperature for each catalyst (A1~B7) is shown as follows:



### REFERENCES

- [1] Liu Yancong, LI Na, LU Jiangyin. Preparation of low carbon olefin from ethanol over P/HZSM-5 Zeolite catalysts [J]. Natural gas chemical industry (C1 chemistry

& chemical industry), 2011, 36(02):32-35.

[2] LU SHAopei. Synthesis of butanol and C<sub>4</sub> olefin by ethanol coupling [D]. Dalian University of Technology, 2018.

[3] Zhang Y F. Improved dynamic cubic exponential

smoothing method for power generation prediction in thermal power plants [D]. Hebei University of Engineering, 2020.

# Airport Taxi Driver Decision-Making and Taxi Management Model

*Renfang Wang<sup>1, 4,\*</sup>, Xueying Jing<sup>1, 4</sup>, Zehui Dou<sup>1, 2, 3</sup>*

<sup>1</sup>*Mathematical Modeling Innovation Lab, North China University of Science and Technology, Tangshan 063210, China;*

<sup>2</sup>*Engineering Computing and Simulation Innovation Lab, North China University of Science and Technology, Tangshan 063210, China;*

<sup>3</sup>*College of Artificial Intelligence, North China University of Science and Technology, Tangshan 063210, China;*

<sup>4</sup>*School of Science, North China University of Science and Technology, Tangshan 063210, China;*

*\*Corresponding Author.*

**Abstract:** Renting a car is a means of transportation that saves social resources and is extremely convenient for people's life. In order to maximize profits, airport taxi drivers often face two choices. Based on this, this paper puts forward some decision-making suggestions. First, taxi drivers can choose to pick up passengers at the airport queuing area, or go directly to the city to pick up passengers. Thus, determine the factors that affect the two choices, contains the paying passengers, airport traffic, the number of the taxi, airport and the distance of the city, taxi price, fuel, etc., and the problem that the taxi driver priority is to maximize profits, so discuss respectively the taxi driver under the two options of revenue function, the driver follow the principle of great benefits to the final decision, The final return is  $M_{total} = \max \{M_1, M_2\}$ . After that, Shanghai Pudong Airport was selected as the research site and relevant data were collected. The monthly passenger throughput of Shanghai Pudong Airport in 2018 was analyzed, and it was found that the passenger throughput in June was closest to the mean value. In order to ensure the universality of the decision, a day in June was selected as the research time. Then, according to the change diagram of one day, it was found that there were significant differences in the number of passengers between daytime and night. In addition, the number of passengers and the number of taxis in question 1 are decision variables. Therefore, the benefits of the two choices in daytime and nighttime are studied respectively. Finally, the concrete decisions under four conditions are obtained.

**Keywords:** Decision model; The maximum income; Selection strategy

## 1. INTRODUCTION

After sending passengers to the airport, taxi drivers often make the decision to empty their vehicles and return to downtown or queue up for passengers to return to downtown according to their personal experience. If too many cars are already parked in the parking lot, the waiting time of taxi drivers will be prolonged. If the number of flights arriving at a certain time is dense, the number of passengers returning to the city by taxi will also increase. Therefore, important influencing factors are selected in the process of establishing the revenue model of airport taxi drivers.

Combining with the specific airport and taxi situation, this

paper gives the taxi driver's decision scheme and discusses the rationality of the model and its dependence on factors. In China, due to the difficulty and inaccuracy of sample data collection of small airports and the large sample data of large airports, it is difficult to conduct statistical analysis. Therefore, Shanghai airport with domestic passenger flow and moderate advanced degree is comprehensively considered to select and analyze taxi drivers' alternative schemes. In the actual situation, taxi drivers will be affected by various determinations and uncertainties, so it is necessary to select the most representative data for analysis. Therefore, in data collection, Shanghai Pudong Airport in 2018, one year before the epidemic, is selected as the research object of the decision-making model., based on the change of the actual monthly and daily traffic in the region, the payment for passenger car, fuel consumption, the average for the driver and the plane arrived at the airport of time interval and a taxi from the parking lot to line up the time interval as the dependent variable yields formula, established for the taxi driver out of options. Finally, the simulation results are compared with the real situation, the rationality of the model is discussed, and the dependence analysis of the influence of some factors on income is conducted.

## 2. EXPERIMENTAL

### 2.1 TAXI DRIVER'S CHOICE DECISION MODEL

Combined with the revenue of airport taxi drivers and the change rule of the number of passengers, the influencing mechanism of factors related to taxi drivers' decision is analyzed and studied, and a choice decision model is established to provide a reasonable choice strategy for taxi drivers.

This paper selects the relevant data of taxi arrivals and departures at the airport in a certain city in China to provide taxi drivers with options, and analyzes the dependence of relevant factors and the rationality of the model.

According to the actual situation, drivers mainly consider the problem of income when carrying passengers, because taxi drivers at the airport are faced with two choices: 1. To the airport queuing area to pick up passengers to the city; 2. Go directly to the city to pick up passengers. Therefore, starting from the income of taxi drivers, the income of the two choices are discussed respectively, and the driver selection decision model is established.

**Step1:** go to the airport queuing area to pick up passengers

to the city. In the waiting area, the total revenue of each vehicle comes from the payment of passengers, minus the revenue of the time spent waiting in the waiting area, that is, the following formula is satisfied:

$$M_1 = m_{\text{pay}} - m_T \quad (1)$$

Wherein, the payment per passenger is the product of the unit price of the taxi and the distance of the passenger, namely:

$$m_{\text{pay}} = m_0 L_1 \quad (2)$$

The revenue of the time spent waiting in the queuing area is the product of the average revenue per minute of each vehicle and the time spent waiting in the queuing area, namely:

$$m_T = T_0 m_1 \quad (3)$$

The waiting time of a taxi in the queuing area consists of two parts: the time interval between the arrival of an airplane at the airport and the time interval between the arrival of a taxi from the parking lot at the queuing area, namely:

$$T_0 = \omega_1 t_1 + t_2 \quad (4)$$

Parameter  $\omega_1$  represents the size of the number of passengers waiting in the queuing area and the number of taxis, namely:

$$\omega_1 = \begin{cases} 0 & \frac{\text{passenger}}{2} > \text{taxis waiting} \\ 1 & \frac{\text{passenger}}{2} \leq \text{taxis waiting} \end{cases} \quad (5)$$

**Step2:** Go directly to urban areas to select passengers. The total revenue of each vehicle is the difference between the revenue of seeing off each guest and the cost of delay when unloaded, i.e., the following formula is satisfied:

$$M_2 = m_{\text{pay}} - m_{\text{empty}} \quad (6)$$

Among them, the cost delayed when there is no load is the fuel cost per minute of the taxi, the product of the distance from the airport to the urban destination and the unit price of the taxi, namely:

$$m_{\text{empty}} = L S m_{\text{single}} \quad (7)$$

**Step3:** When the taxi driver makes a choice, the total revenue meets:

$$M_{\text{total}} = \max\{M_1, M_2\} \quad (8)$$

The final choice of taxi depends on the size of the total revenue of the two ways, namely, the total revenue is the total revenue of carrying passengers from the queuing area to the airport to the urban area and the maximum value of selecting passengers directly to the urban area. Therefore, the final decision model of taxi driver satisfies Equation (8).

## 2.2 THE ESTABLISHMENT AND SOLUTION OF MODEL

### 1. Selection of airports

This question needs to select the taxi data of a specific domestic airport and the corresponding city, select a reasonable plan for taxi drivers according to the obtained model, and verify the dependence of relevant factors. At home because of the small airport sample size is small, the data collection difficult and inaccurate, large sample data is relatively large, at the airport is not convenient to statistics and analysis, therefore, to comprehensively consider the traffic of domestic airports and advanced

level, select the data reliable and easy to handle the Shanghai airport taxi driver selection and analysis of alternatives. Shanghai is a relatively developed city with sufficient taxi flow and sample space.

Figure 1 Shanghai Pudong Airport

Shanghai Pudong International Airport (IATA code: PVG, ICAO code: ZSPD) is a class 4F civil Airport located in Pudong New Area of Shanghai, China. The airport serves 49 international and regional flights to 280 destinations.



Among them, 156 are domestic destinations (including 6 from Hong Kong, Macao and Taiwan) and 124 are international destinations. By 2018, the passenger throughput of Shanghai Pudong Airport reached 74 million, up 5.7 percent year on year. There are rail transit, airport bus, and taxi in the airport. Among them, The development of Shanghai taxi industry is stable and the traffic flow is large, which is very suitable for the study of the passenger selection of airport taxi.

### 2. Conditions and data related to shanghai pudong airport

#### (1) Monthly data changes.

In practice, taxi drivers' decision-making is subject to many deterministic and uncertain factors, such as the impact of the epidemic on airport taxi drivers' decision-making. Therefore, in data collection, Shanghai Pudong Airport in 2018 was selected as the research object to collect the monthly passenger throughput of the airport in 2018, as shown in Table 1 :

Table 1 Monthly throughput and average value of Shanghai Pudong Airport in 2018

time	Passenger	difference	time	Passenger	difference
Jan	288.19	-16.34	Jul	316.56	12.03
Feb	277.88	-26.65	Aug	319.45	14.92
Mar	308.03	3.5	Sep	300.7	-3.83
Apr	318.14	13.61	Oct	324.86	20.33
May	306.42	1.89	Nov	285.35	-19.18
Jun	305.44	0.91	Dec	292.25	-12.28

#### (2) Daily data changes

Taxi drivers are considered in the decision problem in the general case should do what kind of choice, so the number of passengers for 12 months averaging, a value of 3.0453 million people, found in June of passenger throughput and average relatively close, so choose one day in June, data

from the airport as a decision-making data, detailed in table 2:

Table 2 Data changes of Pudong Airport within 24 hours

Time	Passenger	cars	taking taxis	Time	Passenger	cars	taking taxis
1	0.19	400	870	13	0.55	501	828
2	0.15	231	666	14	0.64	472	954
3	0.02	92	97	15	0.87	479	1309
4	0.00	0	0	16	0.75	575	1129
5	0.05	24	75	17	0.70	795	1051
6	0.09	42	131	18	0.67	797	1000
7	0.07	150	107	19	0.67	660	998
8	0.26	201	384	20	0.65	634	972
9	0.16	215	240	21	0.76	590	1147
10	0.67	249	1000	22	0.58	716	868
11	0.84	274	1262	23	0.60	931	906
12	0.52	458	773	24	0.35	734	1570

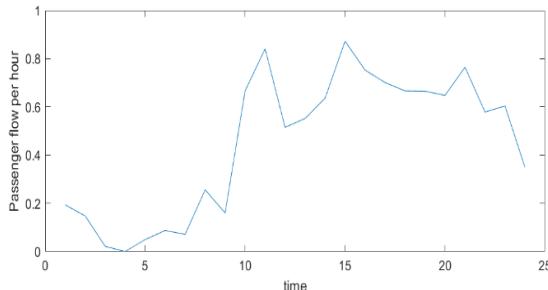


Figure 2 Schematic diagram of one-day passenger flow distribution of passenger flow per hour is shown in Figure 2, and it is found that the passenger flow is the smallest at 4 am and 12 PM. The passenger flow is the largest at 11 am and 3 PM, which is in the peak period of work, and between 10 and 21 PM, except 12 noon, the passenger flow is higher at other times; After 21 o'clock is in a downward trend, 5 to 11 o'clock is in an upward trend.

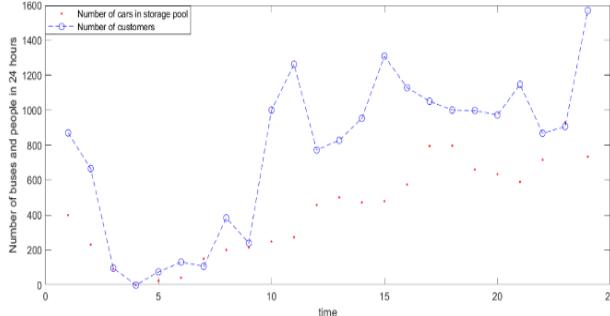


Figure 3 Changes in the number of people in storage pools and taxis

According to Table 2, the changes of the number of cars in the car storage pool and the number of people taking taxis are drawn. As shown in Figure 3, the number of cars in the car storage pool shows an increasing trend from 5 to 23 and a decreasing trend from 23 to 5 in the next day. In addition, the number of people taking taxis shows a decreasing trend from 0 to 5 and an increasing trend from 5 to 24. Therefore, 5-23:00 is the daytime, and 23:00-5:00 is the night, respectively, to discuss the income of daytime and night taxi drivers.

### (3) Guest pays the fee $m_{pay}$

There are differences between daytime and nighttime taxi fare. The conditions of  $m_{pay}$  of daytime taxi fare in Shanghai are as follows: when the distance from the airport to the urban destination is not more than 3 kilometers, the taxi fare is 14 yuan; When it is more than

3 km but less than 10 km, the excess part will be charged 2.5 yuan for each kilometer. When L is greater than 10 km, the excess part will be charged 3.6 yuan per km, that is:

$$\begin{cases} 14 & L \in (0, 3] \\ 14 + 2.5(L - 3) & L \in (3, 10] \\ 31.5 + 3.6 * (L - 10) & L \in [10, \infty) \end{cases} \quad (9)$$

When the distance from the airport to the urban destination is not more than 3 kilometers, the taxi fare is 18 yuan; When it is more than 3 km but less than 10 km, the excess part will be charged 3.1 yuan for each kilometer. When L is greater than 10 km, the excess part will be charged 4.7 yuan per km, that is:

$$\begin{cases} 18 & L \in (0, 3] \\ 18 + 3.1(L - 3) & L \in (3, 10] \\ 39.7 + 4.7 * (L - 10) & L \in [10, \infty) \end{cases} \quad (10)$$

(4) Average revenue per minute  $m_1$ , time interval  $t_1$  of aircraft arriving at the airport.

It is known that the daily operating income of taxis in Shanghai is 1050 yuan, including 350 yuan of miscellaneous fees (taxi repair, etc.) and 250 yuan of fuel for driving 400 kilometers, namely, the total revenue of a day is  $1050 - 350 - 250 = 450$  yuan, and since taxis work 8 hours a day, the average revenue per minute is 0.9 yuan/minute. According to the analysis of all flights in Shanghai, it is found that the table between two flights shall not be less than 13min. In order to ensure that the taxi can receive customers to the maximum extent,  $t_1 = 15\text{min}$  is selected.

(5) Time interval between taxi from parking lot to queuing area  $t_2$ , fuel consumption.

According to the road map, it is about 1km from the taxi parking lot to the queuing area, while the speed limit at this time is 20km/h, so the time interval  $t_2 = \frac{1}{20} * 60 = 3\text{min}$ . It is known that the unit price of gasoline  $m_{dan} = 7.25$  yuan, and the fuel cost for traveling 400 kilometers is 250 yuan, so the fuel consumption is  $S = 250 / 7.25 / 400 = 0.09\text{L/km}$ .

### 2.3 SOLUTION OF TAXI DRIVER DECISION MODEL

To the airport queuing area to pick up passengers to the city

In daytime, the distance from the queuing area to the downtown area is more than 10km, that is, the guests pay the fee and choose the third tier. Based on the known conditions, figure out what the guest should pay[2]:

$$m_{pay} = 14 + 2.5 * 7 + 3.6 * (L - 10) \quad (11)$$

Delay cost:

$$m_T = (\omega_1 * 15 + 3) * 0.9 \quad (12)$$

The final total revenue and the cost payable by the guest and the cost of delay shall meet:

$$M_1 = m_{pay} - m_T \quad (13)$$

Thus, the total daytime return is

$$M_1 = 3.6L - 13.5\omega_2 - 7.2 \quad (14)$$

Similarly, at night, the distance from the queuing area to the downtown area is more than 10km, that is, the guests pay the fee and choose the third gear. Figure out what the guest should pay:

$$m_{pay} = 18 + 3.1 * 7 + 4.7 * (L - 10) \quad (15)$$

Delayed expenses:

$$m_T = (\omega_1 * 15 + 3) * 0.9 \quad (16)$$

The final total revenue and the cost payable by the guest and the cost of delay still meet Figure 4. Thus, the total nightly revenue is

$$M_1 = 4.7L - 13.5\omega_2 - 10 \quad (17)$$

Go directly to the city to pick up passengers.

In daytime, the distance from direct to downtown is more than 10km, that is, the guests pay the fee and choose the third tier. Figure out what the guest should pay:

$$m_{pay} = 14 + 2.5 * 7 + 3.6 * (L - 10) \quad (18)$$

Cost when no load:

$$m_{empty} = L * 0.09 * 7.25 \quad (19)$$

The final total revenue and the fee payable by the guest and the fee when no load:

$$M_2 = m_{pay} - m_{empty} \quad (20)$$

Thus, the total daytime return is

$$M_2 = 2.95L - 4.5 \quad (21)$$

Similarly, the total night cost is

$$M_2 = 4.05L - 7.3 \quad (22)$$

## 2.4 RESULTS

Considering the two choices of taxi drivers and the situation of daytime and night [5] comprehensively, the total income is obtained as follows:

$$\text{Daytime: } \begin{cases} M_1 = 3.6L - 13.5\omega_2 - 7.2 \\ M_2 = 2.95L - 4.5 \end{cases} \quad (23)$$

$$\text{The night: } \begin{cases} M_1 = 4.7L - 13.5\omega_2 - 10 \\ M_2 = 4.05L - 7.3 \end{cases} \quad (24)$$

In the above formula, when in the daytime, if  $\frac{n_{people}}{2} > n_{car}$ , no matter how big is the distance L of the guest from the airport to the urban destination, it satisfies  $M_1 > M_2$ , The choice of  $M_1$ ; where, when  $L = 4km$ , we get  $M_1 = M_2$ , Both can be; If  $\frac{n_{people}}{2} < n_{car}$ , when the distance L from the airport to the urban destination is  $L < 25km$ , find  $M_1 < M_2$ , Select the  $M_2$ , when  $L > 25km$ , find  $M_1 > M_2$ , The choice of  $M_1$ ; Similarly, when at night, if  $\frac{n_{people}}{2} > n_{car}$ , when  $L < 4km$ , get  $M_1 < M_2$ , Select the  $M_2$ , when  $L > 4km$ , get  $M_1 > M_2$ , The choice of  $M_1$ ; If

$\frac{n_{people}}{2} < n_{car}$ , Distance of guest from airport to downtown destination find  $M_1 < M_2$ , Select the  $M_2$ , when  $L < 25km$ ; when  $L > 25km$ , find  $M_1 > M_2$ , The choice of  $M_1$ .

The model is based on the airport taxi service problems in planning process, can be used in other types of large transport hub at the same time, because they belong to the same kind of problem, difference as long as the traffic and traffic hub itself and some parameters, such as a large train stations and hohai port, etc., can be applied to this model solve the problem of passenger station of transhipment.

In practice, there are a lot of certainty and uncertainty factors influence the decision, make the taxi driver to judge correctly, so the airport taxi management according to the actual situation of reasonable "quantitative" batch release a taxi into the "bus" area and arrange a certain number of passengers, to provide reasonable selection strategy for taxi drivers, It is of substantial significance to increase and balance the income of taxi drivers.

## 3.MODEL EVALUATION

### 3.1 ADVANTAGES OF MODEL

Considering a variety of deterministic and random factors that affect drivers' decision-making, the model is put in the position of the driver to think about the problem and establish the model from the perspective of the driver's personal thoughts, and at the same time, the selection mode of the driver is directly explained from the economic level, which is highly practical and reliable.

Analyze and discuss the driver's decision-making strategy on various environmental conditions, and verify the rationality and reliability of the model.

Through specific and reasonable determination of airport taxi pick-up mechanism, the taking efficiency under the number of pick-up points is calculated, and the influence of various environmental parameters on the number of pick-up points is analyzed and discussed, which has suggestions for the overall arrangement of airport management department.

### 3.2 MODEL SHORTCOMINGS

In the process of establishing the model, the time spent in some connection processes, such as the time from the starting point to the storage pool, was ignored, which had a slight impact on the results.

## REFERENCES

- [1] Wei Zhonghua, Wang Lin, Qiu Shi. Optimization of taxi pick-up service station in hub based on queuing theory [J]. Highway traffic science and technology (application technology edition), 2017, 13(10):298-300.
- [2] Su Jing, LIU Yiming. Journal of Science and Technology Economics, 2012, 9(05):202-203.

# Air-Conditioning Model for Car Sharing Based On Kmeans-Dijkstra

Renzhuo Wang<sup>1, 2</sup>, Yi Zhu<sup>1, 3</sup>, Yibo Wang<sup>1, 3</sup>

<sup>1</sup> Mathematical Modeling Innovation Lab, North China University of Science and Technology, Tangshan 063210, China;

<sup>2</sup> School of Artificial Intelligence, North China University of Science and Technology, Tangshan 063210, China;

<sup>3</sup> Yi Sheng College, North China University of Science and Technology, Tangshan 063210, China

**Abstract:** With the development of shared cars, in order to solve the problems of a single shared car model and high operating costs, and increase corporate profitability, a Kmeans-Dijkstra-based air conditioning model for shared cars is proposed. The kmeans clustering algorithm is used to divide car parking spots into four types of shared car demand areas, and analyze the use of shared cars in each area from the perspective of time and space. Based on this, with regional shared car utilization rate and car dispatching routes as constraints, the company's profit amount is the objective function to establish a linear programming model to guide the hourly air conditioning of shared cars. Among them, the Dijkstra algorithm is used to solve the distance between the cluster centers and plan the optimal vehicle dispatching route. It is of great significance to increase corporate profitability.

**Key words:** Dijkstra; K-means clustering; Car sharing; hourly air conditioning

## 1. INTRODUCTION

As an important form of the sharing economy, car-sharing has been closely watched by people and is regarded as one of the important trends in the development of the automobile industry in the future. Since 2015, the car-sharing industry has been "a hundred flowers blooming", and many projects have received huge financing. However, due to problems such as overweight models, high operating costs, and inability to make profits, car-sharing companies have successively closed down because of the broken capital chain. However, due to problems such as overweight models, high operating costs, and inability to make profits, car-sharing companies have successively closed down because of the broken capital chain.

Consumers have the desire to use cars and the desire to upgrade their consumption, but companies have not stopped, constantly seeking lighter and better ways to use cars. The "time-sharing lease" model of shared cars has solved this problem a lot [1]. However, there are too many cost control links in this way, which makes profit very difficult [2]. Looking at the global development of shared cars, there is no very successful operation model and management example [3]. However, the market demand for shared cars is real, and its development still needs to continue to explore new directions.

For this reason, in order to solve the problem that the operation and scheduling mode of shared cars is too single, and to make reasonable use of idle resources, reduce operating costs, and increase business turnover, a local Kmeans-Dijkstra model of air conditioning in shared cars

is proposed. It is of great significance to the formulation of shared car scheduling strategies.

## 2. BASIC ALGORITHM PRINCIPLE

### 2.1 K-MEANS CLUSTERING ALGORITHM

The elbow-shaped graph can roughly estimate the optimal number of clusters. This paper uses the degree of distortion to choose the optimal number of clusters [4].

Define the degree of distortion of each class = the sum of the squares of the distance between the center of gravity of the class and the position of its internal members.

Suppose that  $n$  samples are divided into classes ( $2 \leq K \leq n-1$ , that is, there are two elements in at least one class, and there are at most  $n-1$  elements), denote the  $k$ -th class by  $C_k$  ( $k=1,2,\dots,K$ ), and mark the position of the center of gravity of this class as  $u_k$

The degree of distortion of the class is:

$$\sum_{i \in C_k} |x_i - u_k|^2 \quad (1)$$

Define the total distortion degree of all classes:

$$J = \sum_{K=1}^K \sum_{i \in C_k} |x_i - u_k|^2 \quad (2)$$

The K-means clustering algorithm uses the Euclidean distance to measure the distance between the sample data. First, first randomize the cluster center points, calculate the Euclidean distance of the remaining sample data points to the cluster center, and assign this object to the closest distance In the center cluster, calculate the average value of the data objects in each cluster as the new cluster center, and iterate until the position of the cluster center no longer changes [5].

### 2.2 DIJKSTRA ALGORITHM

Dijkstra algorithm is an effective algorithm for solving the shortest path of a single source point. The breadth-first search is used to relax the edges to solve the single source shortest path in a directed graph or an undirected graph. The algorithm steps are as follows [6-8]:

Step1: Divide all vertices into two parts, the vertex set  $P$  of the known shortest path and the vertex set  $Q$  of the unknown shortest path. The set  $D$  represents the minimum distance between two vertices, and  $d_{ij}$  represents the shortest distance from vertex  $i$  to vertex  $j$

Initialization:  $P$  set only includes source point  $p$ ;  $Q$  set includes all vertices except source point  $p$ . The distance

between two vertices that are not directly connected is infinite.

Step2: In the set  $Q$ , select a vertex closest to the source point and add it to the set  $P$ , and update the minimum distance  $D$  set. Get the current minimum distance  $d_{pj}$  from the source point to the remaining vertices

$$d_{pj} = \min\{d_{pj}\} \quad (3)$$

Step3: Repeat Step2 until the set  $Q$  is empty. At this time, the updated  $d_{pj}$  is the shortest distance from the source point  $P$  to the remaining vertices  $j$ .

### 3. SPATIAL AND TEMPORAL DISTRIBUTION CHARACTERISTICS OF SHARED CARS

In order to rationally mobilize the existing car resources, this paper combines the data given in the certification cup modeling contest and analyzes the temporal and spatial distribution characteristics of shared cars based on the K-means clustering results. The overall flow chart is shown in Figure 1 below:

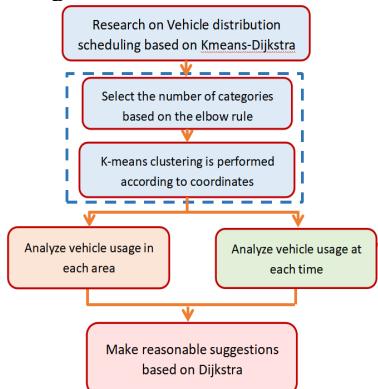


Figure 1 Algorithm flow chart

#### 3.1 DIVISION OF PARKING SPOTS BASED ON K-MEANS

Calculate the aggregation coefficient and solve the reasonable cluster  $K$ . According to the aggregation coefficient line graph, when the number of categories is 4, the downward trend of the line will slow down, so the number of categories can be set to 4. It can be seen from the figure that the distortion degree changes the most when the  $K$  value is from 1 to 4. After more than 4, the degree of distortion is significantly reduced, so the elbow is  $K=4$ , so the number of categories can be set to 4. Where

$J$  is the aggregation coefficient. As shown in Figure 2 below:

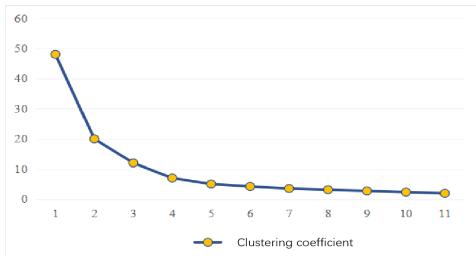


Figure 2 Clustering coefficient graph

According to this, perform K-means clustering:

ACADEMIC PUBLISHING HOUSE

Step 1: Calculate the distance between samples

First initialize the  $k$  initial cluster centers, and then calculate the distance between each point and the cluster center. In this paper, the Euclidean distance is used to calculate the distance between samples. The expression of Euclidean distance is:

$$d(\vec{x}_1 \vec{x}_2) = \sum_{k=1}^p |x_{1k} - x_{2k}| \quad (4)$$

Construct a sample distance matrix of  $1048575 \times 1048575$ :

$$D_0 = \begin{bmatrix} 0 & & & G_1 \\ G_1 G_2 & 0 & & G_2 \\ G_1 G_3 & G_2 G_3 & 0 & G_3 \\ \vdots & \cdots & \cdots & \vdots \\ G_1 G_{1048575} & G_2 G_{1048575} & \cdots & G_{1048574} G_{1048575} & 0 \end{bmatrix} G_{1048575} \quad (5)$$

Step 2: Use the latitude and longitude coordinates of each store to classify the closer points into one category.

Regard each sample as a class, namely  $G_1, G_2, G_3 \dots G_{1048575}$ , observe that the two classes  $G_i$  and  $G_j$  with the smallest Euclidean distance

are grouped into one class, denoted as  $G_{1048576}$ , and calculate the distance between the new class and the other classes except  $G_i$  and  $G_j$ , Get a new distance matrix  $D_i$ .

$$D(G_{1048576}, G_1) = \min\{D(G_i, G_1), D(G_j, G_1)\}$$

$$D(G_{1048576}, G_1) = \min\{D(G_i, G_2), D(G_j, G_2)\} \quad (6)$$

⋮

$$D(G_{1048576}, G_1) = \min\{D(G_i, G_{1048576}), D(G_j, G_{1048576})\}$$

Update the  $1048574 \times 1048574$  distance matrix  $d$  according to the calculated Euclidean distance between classes  $D_1$

The vehicle information of longitude and latitude is selected to classify according to the optimal number of categories obtained by the elbow rule, and 4 kinds of distribution clustering information scatter plots are obtained, as shown in Figure 3.

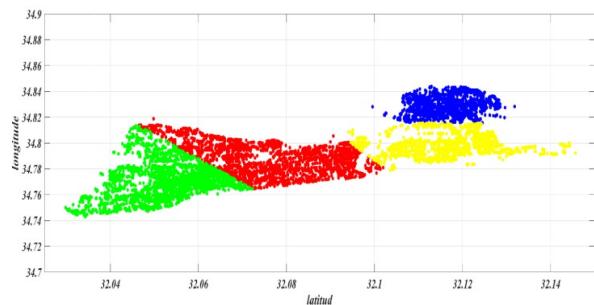


Figure 3 Clustering figure

From bottom to top, the four cluster centers correspond to the vicinity of the tourist attraction Suspended Orange Tree; the vicinity of Dubonov Park; the vicinity of Ramat Aviv Mall; and the vicinity of Revivim Road. According to the relevant information, these 4 locations are all densely distributed locations, so it is reasonable to divide

all vehicle distribution areas into 4 categories and analyze them separately.

### 3.2 SPATIAL DISTRIBUTION OF PARKING SPOTS

Based on the K-means clustering results and the heat map of the corresponding area, the team can preliminarily determine the approximate distribution of vehicles in different areas. Based on the K-means clustering results and the heat map of the corresponding area, the team can preliminarily determine the approximate distribution of vehicles in different areas.

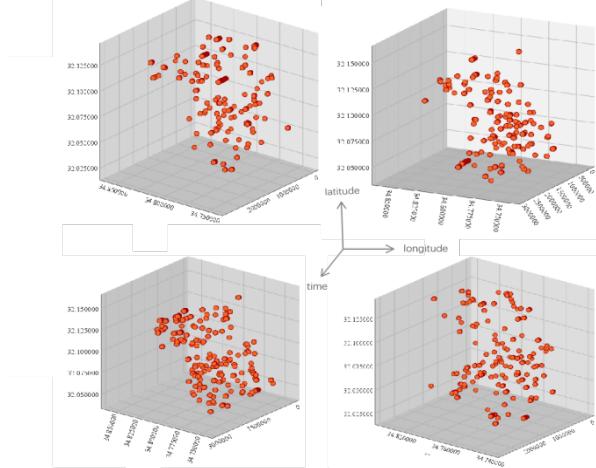


Figure 4 Scatter diagram of vehicle changes over time in category 4

The three-dimensional scatter chart shown in the figure above can very intuitively show the changes in the spatial position of shared cars over time, and the high-frequency overlap at a fixed latitude and longitude position, and can judge the size of the demand for shared cars.

From the scatter diagram, it can be seen that for some areas, the scatter diagram shows a high degree of overlap. In reality, it shows that some shared cars are parked in certain areas for a long time without moving. Therefore, it can show that the area is more effective for shared cars. The demand for shared cars is small; on the contrary, the scatter plots in some areas are denser but do not show overlap. It shows that shared cars often appear in different locations in this area. Therefore, it can be judged that the demand for shared cars in this area is relatively high. After classifying and summarizing the latitude and longitude changes of the shared car samples studied, the team screened out the three longitude and latitude coordinates with the longest parking time for shared cars in each category, as shown in Table 1 below:

Table 1 The time and coordinates of vehicles parked for longer periods in each class

latitude	longitude	Parking time (seconds)
32.12666	34.80831	49014
32.12673	34.8083	49283
32.09674	34.80412	87170
32.05178	34.805	68362
32.05177	34.80497	25129
32.12612	34.83551	71474
32.12616	34.83523	74536
32.11831	34.84284	29255

232.11285	34.82127	41517
32.058374	34.77573	64165
32.0421	34.77948	56663
32.04612	34.80288	58745

### 3.3 TIME DISTRIBUTION OF PARKING SPOTS

It is considered that the data with the same license plate number at different latitudes and longitudes can indicate the flow of the vehicle. If the vehicle is at the same latitude and longitude position, it means that the vehicle has not moved. Therefore, only the data with the closest date to today is the final parking time of the vehicle. Use Excel and Python to perform related processing, compress the time of the vehicle that is at the same latitude and longitude but the time has changed into the time at the last position, and finally obtain the time data sample of each license plate number at different latitude and longitude, and use this as an analysis The basis of vehicle usage in each time period.

Draw the frequency distribution of the above-mentioned processed vehicle flow in different regions and different time periods into a frequency distribution diagram, as shown in Figure 5 below:

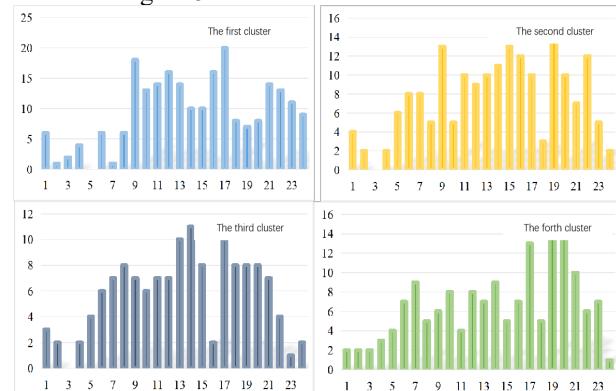


Figure 5 Vehicle flow in different areas at different time periods

The abscissa in the figure represents each time period in a day, and the ordinate represents the review of vehicle flow. From the graph analysis, we can see that the peaks and valleys of local vehicle flow corresponding to different regions are different. For example, the morning and evening peaks of shared car usage near Ramat Aviv Mall in the first type of area are 9 am and 5 pm; the second type of area represents the peak of vehicle use near the Bloodfield Stadium at 9 am and afternoon respectively. 3 o'clock and 7 o'clock in the evening, and at a time different from the first type of areas, there are fewer vehicles from 5 o'clock in the evening to 6 o'clock in the evening; The peak hours of vehicle usage in the areas near Revivim Street represented by the third type of area are 2 pm and 5 pm respectively; the peak hours of vehicles near Dubnov Park represented by the fourth type of area are 5 pm and 7 pm to 8 pm.

From an overall point of view, the overall shared vehicle usage time in the area ushered in a peak from 8 to 9 in the morning, and then declined; it will usher in a peak again from 13 to 14 in the afternoon, and the use of vehicles in the area will usher in from 15 to 17 o'clock. Come to the

trough, and then there will be a peak of vehicle usage from 19 to 20 o'clock.

Due to the different functions corresponding to different regions, there will be differences in the time period of shared car usage between regions and regions. The representative areas are the area around Ramat Aviv Mall represented by the first type of area and the second type of area. In the area near the Bloodfield Stadium, the two places ushered in the peak of car use at 9 a.m., and the use of shared cars from 11 to 3 p.m. declined. However, the second type of area was between 7 p.m. and 8 p.m. The frequency of car use is higher than in the first category. Due to the different functions corresponding to different regions, there will be differences in the time period of shared car usage between regions and regions. The representative areas are the area around Ramat Aviv Mall represented by the first type of area and the second type of area. In the area near the Bloodfield Stadium, the two places ushered in the peak of car use at 9 a.m., and the use of shared cars from 11 to 3 p.m. declined. However, the second type of area was between 7 p.m. and 8 p.m. The frequency of car use is higher than in the first category.

#### 4. BASED ON DIJKSTRA'S AIR-CONDITIONING SCHEME BETWEEN PARKING SPOTS

Based on K-means clustering, the position and coordinates of the cluster center points are obtained as shown in Table 2 below:

Table2 Coordinates of Cluster Center points

Cluster center point	longitude	latitude
Suspended Orange Tree	34.79723	32.11364
Dubonov Park	34.76511	32.05107
Ramat Aviv Mall	34.83206	32.11597
Revivm Road	34.78412	32.07666

Using these four clustering center points as the node positions of shared car scheduling, according to the geometric knowledge of the sphere and the Pythagorean theorem, the arc length between any two points  $A$  and  $B$  can be obtained as:

$$\hat{AB} = R \times \arccos(\cos \theta_i \cos \theta_j \cos(\varphi_i - \varphi_j) + \sin \theta_i \sin \theta_j) \quad (7)$$

Where  $R$  is the radius of the earth,  $\theta_i$ ,  $\theta_j$  are Latitude of point A, B,  $\varphi_i$ ,  $\varphi_j$  are Longitude of point a, b.  $A$ ,  $B$ .

Therefore, the distance between each cluster center node can be calculated as shown in Figure 6 below;



Figure 6 Distance chromaticity

Consider the company's profitability requirements, select the optimal car scheduling plan, and construct a linear objective function, that is, take the company's maximum profit as the objective function, and the profit is the rental car rental minus the cost of car purchase, and the cost of vehicle scheduling and transportation.

$$\max W = \sum_{i=1}^4 \sum_{j=1}^4 (\lambda - d_{ij} * \omega_d - \alpha) * N_{ij} \quad (8)$$

Among them,  $\lambda$  is the average user renting a car and paying the fee,  $d_{ij}$  is the distance from area  $i$  to area  $j$ ,  $\omega_d$  is the transportation cost per kilometer for each shared car,  $\alpha$  is the unit price of each shared car, and  $N_{ij}$  is from the area The number of vehicles dispatched from  $i$  to area  $j$ .

Calculate the weight of the route between the two points according to the profit of the objective function. Among them, according to the actual traffic route and traffic conditions, the direct path between Suspended Orange Tree and Dubonov Park is discarded, and the direct path between Ramat Aviv Mall and Dubonov Park is stored in the form of an adjacency matrix, which is a matrix, By adopting Dijkstra algorithm, the optimal route for car scheduling between any cluster centers is obtained.

#### 5. CONCLUSION

This paper divides shared cars into 4 clusters through K-means clustering, analyzes the temporal and spatial distribution of the four regions, uses Dijkstra algorithm, combined with the most profitable objective function, and finds the optimal route for car dispatch between regions. And put forward reasonable suggestions.

(1) The areas near the Schuster Center, near the entrance to the sea, major hotels in the south, and nearby government agencies are areas with high demand for shared cars, and the initial locations of shared cars are scattered, so shared cars can be placed as much as possible. The above area.

(2) The number of shared cars distributed in the west and south of Tel Aviv Central Railway Station is large and scattered, but the demand for shared cars in this area is not high, and the shared cars in this area can be put into high demand Area.

(3) In some areas, such as the historic landmarks of Afeka Caves, Darom Park, and Hayarkon Park Lake, although the flow of people may be dense, the demand for shared vehicles is not large, so the number of cars placed in the area can be reduced.

(4) Strengthen the supervision of shared cars to avoid long-term parking of some cars in places where the demand for shared cars is small and the flow of people is small, such as Kadmoni Asa Street, Levinski Street, etc.

(5) Make full use of the flexibility of shared car scheduling in time and space, give different scheduling plans for different regions and different time periods, fully consider the functions of each region, and dispatch

vehicles in the region according to their functions, thereby reducing costs, Expand income.

(6) From an overall point of view, the peak vehicle usage in this area is from 8 to 9 in the morning, 13 to 14 in the afternoon, and 19 to 20 in the evening. The company should reasonably consider the time period when consumers use the vehicle, and use the time period when the vehicle is less used to do maintenance and other work for the shared car.

#### REFERENCES

- [1] Xu Yizhan, Liu Xiaoyu, Mo Wentao, Wang Baixuan, Zhang Haixia. Research on the New "Car Sharing" Business Model[J]. China Business Journal, 2021(19):132-135.
- [2] Chen Bo. Research on my country's shared car development strategy based on SWOT analysis[J].Journal of Chongqing Jiaotong University (Social Science Edition), 2020, 20(04):41-46.
- [3] Wang Jing. The global shared car industry still faces multiple obstacles[N]. Economic Information Daily, 2021-03-26 (006).
- [4] Jia Ruiyu, Li Yugong.K-means algorithm of self-determining the number of clusters and initial center point[J]. Computer Engineering and Applications, 2018, 54(07):152-158.
- [5] Yang Junchuang, Zhao Chao. Summary of K-Means clustering algorithm research[J]. Computer Engineering and Applications, 2019, 55(23):7-14+63.
- [6] Xu Weijie, Gao Hairuo.Multi-source optimal escape route planning based on improved Dijkstra algorithm[J]. Communication World, 2019, 26(05):241-243.
- [7] Cao Shuhuai, Wang Xiao, Jiang Haoran, Liang Xiao, Qu Fang. The application of improved Dijkstra algorithm in the optimal path problem of emergency rescue[J]. Shandong Industrial Technology, 2017(01):144.
- [8] Sheng Xin, Chen Peishuai.Research on the shortest path algorithm (dijkstra algorithm) [J]. Science and Technology Information, 2008(26):82-84+44.

# Credit Decision Optimization Model Based on Linear Programming

Sunyufei Wang<sup>1,2</sup>

<sup>1</sup>Mathematical Modeling Innovation Lab, North China University of Science and Technology, Tangshan 063210, China;

<sup>2</sup>School of Artificial Intelligence, North China University of Science and Technology, Tangshan 063210, China

**Abstract:** Credit risk assessment and decision optimization problem is a focus in the study of bank credit. Aiming at this problem, this paper has established the two mathematical models One for the enterprise credit risk assessment based on entropy weight TOPSIS evaluation model, another for bank credit decision planning optimization model based on linear programming First of all, the credit problem is divided into two parts in this paper Credit risk assessment and credit strategy optimization Then, the data preprocessing for Banks, that is, remove the invalid invoice in enterprise's profit income growth rate and the annual rate of growth of output of the upstream company number number downstream company traded with the upstream company and deal with the downstream company number Finally, training the model with data processed.

**Keywords:** Entropy weight method; TOPSIS; multiple linear regression; linear programming credit; decision making

## 1. INTRODUCTION

Financial innovation and the change of personal consumption concept promote the rapid development of domestic bank credit business, especially personal credit business. The scale enlargement of loan business at the same time, however, also brought huge risks to the bank, if borrowers unable or unwilling to repay principal and interest according to the credit agreement, there will be the loss of bank management, how to expand at the same time, effective control and guard against personal credit risk is an important problem of the big Banks need to solve. Therefore, the construction of personal credit risk assessment system, timely detection of dangerous loan applications, effectively deal with the possible personal credit risks, both for the commercial banks themselves, as well as for the regulatory authorities, has important practical significance [1].

Recently, the Central bank announced a comprehensive RRR reduction, one of its intentions is to guide financial institutions to actively use the RRR reduction funds to increase the support for small and micro enterprises, which shows the central government attaches importance to small, medium and micro enterprises. The biggest problem of financial support for micro, small and medium-sized enterprises is that the risk is difficult to determine, and risk management is the focus of the financial industry. At present, the credit risk of micro, small and medium-sized enterprises is determined by the transaction bill information and the influence of upstream and downstream enterprises. Therefore, at present, a bank

credit risk assessment model for small and medium-sized enterprises is particularly important.

## 2. DATA PROCESSING

### 2.1 DATA ACQUISITION

First of all, the financial data of previous years disclosed by the bank was obtained through the web crawler, and the information of input invoice and output invoice was deleted as invalid invoice. Then calculate the profit of each enterprise, the annual growth rate of input and output, the number of upstream companies, the number of transactions with upstream companies, the number of downstream companies and the number of transactions with downstream companies, the corresponding data of seven indicators. Finally, the enterprise strength evaluation data set is obtained.

Each enterprise's profit, annual growth rate of income and annual growth rate of sales are used to describe the profitability of the enterprise, and the number of upstream companies, the number of transactions with upstream companies, the number of downstream companies and the number of transactions with downstream companies are used to describe the influence of upstream and downstream companies.

Finally, the data set obtained is shown in Table 1.

Table 1. Enterprise evaluation data set

Profits	Entry growth rate	Sale growth rate
-1939308588	5.443962412	1.056589537
463807677.3	1.648907527	2.336769368

Upstream companies Number	Downstream	Upstream times	Downstream times
436	352	3249	7886
3625	1579	31435	11665

### 2.2 DATA STANDARDS AND WEIGHT PROCESSING

The dimension of the data in this paper has a great impact and should be standardized first. The annual growth rate of sales items is distributed around 1, while the total number of downstream is distributed around 1000 and up to more than 20000. Therefore, it can be seen that dimensionality exists between data, which has a great influence on the results.

In this paper, the following formula is used to standardize the values of each indicator:

$$Y_{ij} = \frac{X_{ij} - \min(X_i)}{\max(X_i) - \min(X_i)}$$

Information entropy is the probability of occurrence of discrete random events. Its calculation formula is as follows:

$$E_j = -\ln(n)^{-1} \sum_{i=1}^n p_{ij} \ln p_{ij}$$

Where,  $p_{ij} = Y_{ij} / \sum_{i=1}^n Y_i$ , if  $p_{ij} = 0$ ,  $\lim_{p_{ij} \rightarrow 0} p_{iv} \ln p_{ij} = 0$  is defined.

The weight of each index is calculated by information entropy, and the formula is as follows:

$$W_i = \frac{1 - \varepsilon_i}{k - \sum E_i} \quad (i = 1, 2, \dots, k)$$

### 2.3 TOPISS SCORE

Define the distance between the  $i$  th evaluation index and the maximum value as follow:

$$D_i^+ = \sqrt{\sum_{j=1}^m \omega_j (Z_j^+ - z_{ij})^2}$$

Define the distance between the  $i$  th evaluation index and the minimum value as follow:

$$D_i^- = \sqrt{\sum_{j=1}^m \omega_j (Z_j^- - z_{ij})^2}$$

Therefore, the normalization score is as follow:

$$S_i = \frac{D_i^-}{D_i^+ + D_i^-}$$

## 3. EVALUATION MODEL BASED ON MULTIPLE LINEAR REGRESSION

### 3.1 ESTABLISHMENT OF EVALUATION MODEL

In this paper, multiple linear regression is adopted to establish an evaluation model to predict each index coefficient.  $f_1, f_2, f_3, f_4, f_5, f_6, f_7$  is the independent variable,  $y$  is the dependent variable, and the function of the model is expressed as follows:

$$\beta_0, \beta_1, \dots, \beta_p = \arg \min_{\beta_0, \beta_1, \dots, \beta_p} \left( \sum_{i=1}^n (y_i - \hat{y}_i)^2 \right) = \arg \min_{\beta_0, \beta_1, \dots, \beta_p} \left( \sum_{i=1}^n \left( y_i - \beta_0 - \sum_j \beta_j f_{ij} \right)^2 \right)$$

$$y_i = \beta_0 + \sum_j \beta_j f_{ij} + u_i, i = 1, 2, \dots, n, j = 1, 2, \dots, P$$

### 3.2 QUANTITATIVE STANDARDS

Based on the above scores, this paper formulated the following quantitative standards. Grades A, B, C and D are 4, 3, 2 and 1, respectively.

Table 2. Table of significance test results

Source	SS	df	MS	Number of obs=123 F(7, 115)=1598.19
Model	0.008926594	7	0.001275228	Prob>F=0.0000
Residual	0.000091761	115	7.9792e-07	R-squared=0.9898
Total	0.009018355	122	0.000073921	Root MSE=0.00089

It can be seen from the above table that  $0.008 < 0.05$  corresponds to F test, that is, the regression equation is considered meaningful within 95% confidence interval.

Where,  $R^2$  and  $R^2_{adj}$  are close to 1, indicating that the

$$\begin{cases} 4, & y \geq 0.04 \\ 3, & 0.02 \leq y \leq 0.04 \\ 2, & 0.008 \leq y \leq 0.04 \\ 1, & y \leq 0.008 \end{cases}$$

### 3.3 OPTIMIZATION OF THE MODEL

In this paper, the linear programming algorithm is used to optimize the credit strategy. The linear programming model is established under the constraint conditions that the maximum bank income is the objective function, the credit limit of a single enterprise is 1~100 million yuan, the annual interest rate is 4~15%, and the total amount of enterprise loans is constant.

$$\max Z \quad s.t. 10 \leq b \leq 100 \quad 4\% \leq c \leq 15\% \quad \sum_{i=1}^n A_i = C$$

Where,  $Z$  is bank income,  $S$  is bank risk, and  $C$  is constant.

$$Z = \sum_{i=1}^3 A_i (1 - I_i) r_i$$

Where,  $A_i$  is the amount of bank loans to enterprises,  $I_i$  is customer churn rate, and  $r_i$  is the interest rate of enterprise loans.

## 4. MODEL RESULTS AND DISCUSSION

### 4.1 MODEL CALCULATION RESULTS

The score results calculated in this paper are as shown in Figure 1

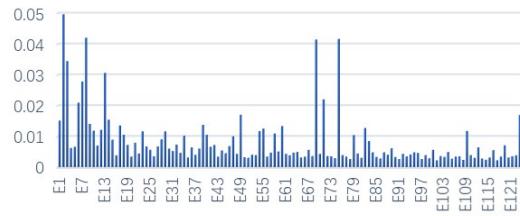


Figure 1. Score bar chart

### 4.2 RESULTS OF SIGNIFICANCE TEST

The results of significance test are as shown in Table 2.

regression model has a high goodness of fit.

### 4.3 SOLUTION OF MULTIPLE LINEAR REGRESSION EQUATION

The regression coefficient is shown in Table 3.

Table 3. Regression coefficient table

score	Coef	Std.Err	t	p> t	[95%conf.Interval]
profit	1.67e-13	3.13e-13	0.53	0.595	-4.53e-13 7.86e-13
Annual growth rate of entries	0.0000386	9.30e-07	41.49	0	0.0000367 0.0000404
Annual growth rate of sales	0.0002094	6.20e-06	33.78	0	0.0001971 0.0002217
Number of upstream companies	5.27e-06	7.02e-07	7.51	0	3.88e-06 6.66e-06
Number of downstream companies	3.90e-06	1.20e-07	32.45	0	3.66e-06 4.13e-06
The total number of upstream times	3.34e-07	8.56e-08	4.00	0	1.73e-06 5.12e-07
The total number of downstream times	1.12e-06	3.37e-08	30.18	0	1.05e-06 1.20e-06
cons	0.0011503	0.0001228	9.37	0	0.0009071 0.0013934

After solving the index coefficient, the regression equation is as follows:

$$y = 1.67e - 13f_1 + 0.0000386f_2 + 0.0002094f_3 + 5.27e - 06f_4 \\ + 3.90e - 06f_5 + 3.34e - 07f_6 + 1.12e - 06f_7 + 0.0011503$$

#### 4.4 Model integration

In this paper, the relationship between the loss rate of different types of enterprises and the interest rate of bank loans is solved by interpolating and fitting methods, and the relationship between bank loan interest rates and annual interest rates is as follows:

$$I_j = a_i \sin(b_i r_i + c_i)$$

The fitted result is shown in Figure 2.

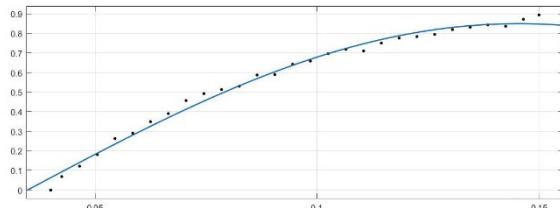


Figure 2. Fitting diagram of the relationship between the attrition rate of Class A enterprises and the interest rate of bank loans

The final result is shown in Table 4.

Table 4. Coefficient table

category	a	b	c
Class A	0.8834	14.89	-0.4991
Class B	0.848	14.49	-0.49451
Class C	0.85	14.19	-0.4934

Since both the interest rate and the loan amount are unknown quantities, the maximum quantity of 1 million yuan will be fixed when solving the interest rate to solve the loan interest rate. The loan interest rates A, B and C are 0.0400.

Put customer churn rate  $I_i$  and corporate loan interest rate

$r_i$  into the target function to obtain the loan amount as shown in Table 5.

Table 5. Table of Loan Amounts

category	Amount
Class A	1000000
Class B	900000
Class C	600000

This paper also attempts to limit the total amount of loans to 100 million yuan for the whole year, so based on the original model, this paper changes the credit strategy for enterprises and introduces a new bank interest function, as follows:

$$S = n_A \square A(1 - a_1 \sin(b_1 r_i - c_1)) + n_B \square A(1 - a_2 \sin(b_2 r_i - c_2)) + n_C \square A(1 - a_3 \sin(b_3 r_i - c_3))$$

And Where  $n_A$ ,  $n_B$ ,  $n_C$  and correspond to the number of enterprises with different credit rating, and the corresponding distribution was 26 in Class A, 141 in class B, 83 in class C and 51 in class D. Figure 3. Distribution map of corporate reputation score

And give the corresponding constraints as follows:

$$n_A A + n_B B + n_C C \in [9900, 10000] \quad A \in [10, 100]$$

$$B \in [10, 90] \quad C \in [10, 60]$$

Finally, through the solution of the other side, it is concluded that for each credit rating enterprise to give different credit lines, the loan amount of Class A

ACADEMIC PUBLISHING HOUSE

enterprises is 600, 000 yuan, the loan amount of Class B enterprises is 335, 000 yuan, and the loan amount of Category C enterprises is 100, 000 yuan. The corresponding lending rate is 4%.

#### 4.4 DISCUSSION OF MODEL RESULTS

Firstly, the model is divided into two parts: credit risk assessment and credit strategy optimization. Then, the data is preprocessed, that is, the invalid invoice is removed, the profit of the enterprise is calculated, the annual growth rate of input and output, the number of upstream companies, the number of transactions with upstream companies, the number of downstream companies and the number of transactions with downstream companies are calculated. Finally, two models are trained by using the processed data.

Credit risk assessment model: First, the entropy weight method is used to calculate the weight of the 7 indicators. Then, a weighted TOPSIS method was used to score each firm. Finally, with 7 indicators as independent variables and scores as dependent variables, A multiple linear regression equation is established, and quantification standards are set to divide enterprises into four levels: A, B, C and D.

Optimization model of credit strategy: firstly, with the maximum bank income as the objective function, a linear programming model is established to calculate the loan interest rate and the loan amount of a single enterprise with the loan amount of 100~1 million yuan, the loan interest rate of 4%~15%, and the total loan amount of constant as the constraint conditions. Then, the loan amount is fixed at the maximum to solve the loan interest rate. Finally, the loan interest rate is substituted into the target function to obtain the loan amount.

The model proposed in this paper analyzes the loan risks of small and medium-sized enterprises from various aspects, and gives classification, so that the bank can have a tailored loan plan for each enterprise, and can accurately control the actual situation of each enterprise, so as to maximize the bank's income.

Between the production and operation of enterprises and economic benefits could be affected by some factors, corresponding to on-time delivery of loan interest, in order to common development of enterprises and the Banks, can consider introducing enterprise anti-risk ability index, to evaluate each enterprise in the face of the sudden and unexpected circumstances, such as new crown outbreak to the enterprise is difficult to produce income situation, Through the study of the enterprise's own capital and daily operating costs, this paper finally obtains the anti-risk capability function of the corresponding enterprise:

$$T = \frac{M_j - N_j}{N_j} D_j$$

Wherein, the annual input amount is  $M_j$ , the annual output amount is  $N_j$ , and the annual consumption date is  $D_j$ .

By inserting the data of each enterprise into the formula, this paper obtains the score of each enterprise's anti-risk

ability, classifies the final score through K-means clustering, and finally obtains the rating standard of enterprise's anti-risk ability, the result is shown in Table 6.

Table 6. Enterprise anti-risk ability score

classification	Enterprise anti-risk ability score
Class A	$1078 \leq T$
Class B	$100 \leq T \leq 1078$
Class C	$0 \leq T \leq 100$
Class D	$T \leq 0$

Finally, preferential loan policies are implemented for category A, B and C enterprises, while for Category D enterprises, considering the loan risks of the bank itself, this paper does not provide loans to Category D enterprises. The corresponding loan policies are shown in Table 7.

Table 7. Loan interest rate after policy change

classification	Loan interest rate based on policy change
Class A	3.5%
Class B	3.0%
Class C	2.5%

For enterprises with poor risk resistance, this paper provides more preferential policies to reduce the loan interest rate, so as to help enterprises tide over difficulties without excessive losses for banks.

#### 4.5 APPLICATION OF THE MODEL

In 2020, the outbreak of COVID-19 hit the whole country. Small, medium and micro enterprises have the characteristics of little self-owned capital, small asset scale, weak risk resistance, short life cycle, relatively single operation, high market elimination rate and high business risk, making it difficult for them to survive in the epidemic environment. Most enterprises did not get enough financial support in time and were forced to close down.

Due to the instability of small, medium and micro enterprises themselves, domestic banks dare not lend, or even refuse to borrow, most enterprises can not get financing. Most domestic credit institutions have less credit capital, extensive regulatory means and imperfect institutional mechanisms, making it more difficult for

small, medium and micro enterprises to survive. Many enterprises have already been "struggling to survive" due to the impact of the epidemic.

However, the credit risk assessment model in this paper is a good solution to the current plight of smes and banks. The model can extract more key credit data features of small and medium-sized enterprises, so as to provide banks with visual data display and correlation analysis, and the results are more consistent with real life, so as to facilitate the subsequent management of banks and ensure the long-term stability of loan business. At the same time, our model also tends to pay more attention to the development of small and medium-sized enterprises, which fills the gap in capital supply and demand structure caused by the characteristics of different industries and prevents banks from investing too much loans into industries already saturated with capital.

Improving the assessment of credit risks of commercial banks can promote the development of credit loans for small and micro enterprises by commercial banks, so as to help small and micro enterprises obtain sufficient financial support from commercial banks. Commercial banks small and micro enterprise credit loan risk assessment system, in the commercial banks small and micro enterprise loan risk assessment application, can play an effective role in stabilizing China's financial environment, promote the development of China's financial cause.[2]

#### REFERENCES

- [1] CAI Wenzhe, LUO Yonghao, ZHANG Guanxiang, ZHONG Huiling. Personal credit risk Assessment Model based on GBDT and Logistic Regression and empirical Analysis [J]. Management modernization, 2017, 37(02):1-4.
- [2] Li Ronghua. Empirical research on credit risk assessment model of small and micro enterprises in commercial banks [J]. Finance and Economics, 2016(16):158-159.

# Analysis and Description of Oxygen Saturation based on Pearson Correlation Coefficient and Multiple Minear Regression

Yunjiang Han<sup>1, 2</sup>, Yibo Wang<sup>2, 3</sup>, Yi Zhu<sup>2, 3</sup>

<sup>1</sup>College of Mechanical and Electrical Engineering, Shenyang Aerospace University, Shenyang 110135, China;

<sup>2</sup>Mathematical Modeling Innovation Lab, North China University of Science and Technology, Tangshan 063210, China;

<sup>3</sup> Yi Sheng College, North China University of Science and Technology, Tangshan 063210, China

**Abstract:** Oxygen saturation is the percentage of the oxyhemoglobin bound by oxygen in the blood hemoglobin volume, which is an important physiological parameter of the respiratory circulation. The functional oxygen saturation is the ratio of HbO<sub>2</sub> concentration to HbO<sub>2</sub> + Hb concentration, which is different from the percentage of oxygenhemoglobin. Therefore, monitoring the arterial oxygen saturation allows for the estimation of the oxygenation and hemoglobin oxygen-carrying capacity of the lungs. Considering the inter-individual variability and the possible interference of individual activity and environmental differences with the experimental data, the noise reduction of the experimental data is required first. After the processing, the obtained optimal data is processed for dimensionality reduction, and the correlation coefficient matrix of the indicators and the original index after the dimension reduction is calculated to obtain the correlation between the indicators, so as to obtain the specific meaning of the indicators after the dimension reduction. Next, a regression analysis was conducted on the indicators after dimension reduction and the mean blood oxygen saturation of the experimental individuals, and the regression results were hypothesis tested to obtain the reasonable regression results satisfying the hypothesis test, so as to comprehensively evaluate the influencing factors of blood oxygen saturation.

**Keyword:** Pearson Correlation Coefficient; Multiple Linear Regression; Principal Component Analysis; Descriptive Statistic

## 1. INTRODUCTION

In order to indicate the origin of Oxygen Saturation, the following background is worth mentioning. Oxygen saturation(SpO<sub>2</sub>) is oxygenated hemoglobin (HbO<sub>2</sub>) in the blood. It is an important physiological parameter of respiratory circulation. And the functional oxygen saturation is HbO<sub>2</sub>Concentration and HbO<sub>2</sub>+Hb concentration ratio is different from the percentage of oxygenated hemoglobin. Accordingly, arterial oxygen saturation (SaO<sub>2</sub>) was monitored. Oxygenation and hemoglobin oxygen-carrying capacity in the lungs can be estimated.

It is an important physiological parameter of respiratory circulation, and monitoring arterial oxygen saturation can evaluate lung oxygenation and hemoglobin oxygen carrying capacity. On the other hand, oxygen saturation can also reflect human health to some extent. The oxygen

saturation of normal human arterial blood is about 98%. At present, photoelectric pulse wave detection method is mainly used to detect blood oxygen saturation.

## 2. THE DESCRIPTION OF THE PROBLEM

Pulse oximetry is routinely used for monitoring patients oxygen saturation levels. During continuous monitoring, maybe there is a way can be able to describe the patterns of oxygen saturation using a model.

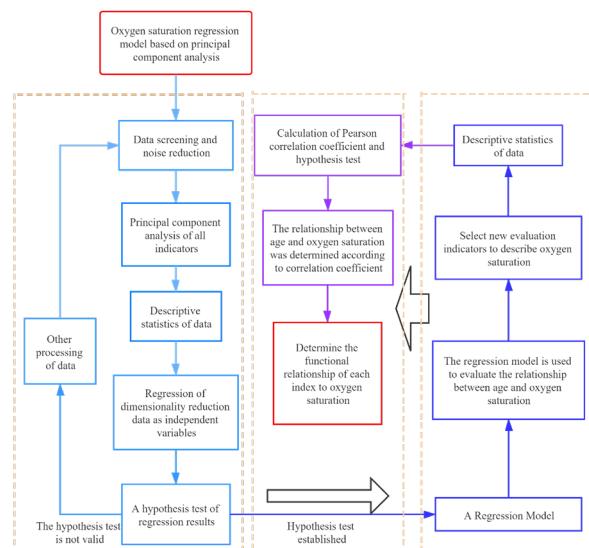
It have the data of 36individuals, each subject was tested the oxygen saturation continuously for approximately 1hour at a frequency of 1Hz. It also recorded the following information about the participants, including age, BMI, gender, Smoking history and current smoking status, and any significant medical conditions that could affect reading.

Now want to use these data to find typical patterns of variability in oxygen saturation so that we could use several parameters to characterize an individual. It would also like to know whether the pattern of oxygen saturation series is related to age, i.e., which characteristics change in older people compared to younger people.

## 3. MODELS

### 3.1 ANALYSIS OF ISSUES

#### 3.1.1 ANALYSIS OF ISSUES



First of all, the differences between individuals and the changes of individual activities and environment may interfere with the experimental data. In order to eliminate or minimize this interference, the experimental data need

to be de-noised before the model is established. Considering the advantages and disadvantages of different data processing methods, a series of methods such as variance, square, logarithm and square are used to process the data. After noise reduction, the meaningful data are PCA dimension reduction, and the correlation coefficient matrix between each index and the original index is calculated to obtain the correlation between the original index and the dimension reduction index.

### 3.1.2 TERMS, DEFINITIONS AND SYMBOLS

Symbols	Implications
y1	Mean Oxygen Saturation
y2	Standard deviation of oxygen saturation
x1	Body Mass Index
x2	Body Health Index
x3	Body Mass Index
x4	The Square of the age
P	Hypothesis test
$\beta$	Correlation

In order to obtain the specific meaning of each index after dimension reduction. The index after dimension reduction

Table I: Descriptive statistical results of indicators

Descriptive statistics	Sex (male)	Sex (female)	Non-smoker	Pre-smoking	Smoking	BMI	age	average	standard deviation
minimum value	0.000	0.000	0.000	0.000	0.000	17.900	19.000	93.748	0.342
maximum value	1.000	1.000	1.000	1.000	1.000	28.400	70.000	99.508	1.292
mean value	0.472	0.528	0.778	0.139	0.083	23.136	33.889	97.787	0.702
median	0.000	1.000	1.000	0.000	0.000	23.400	23.000	98.174	0.688
bias angle	0.111	-0.111	-1.336	2.088	3.015	-0.004	0.755	-1.292	0.790
kurtosis	1.012	1.012	2.786	5.361	10.091	2.191	2.194	5.531	3.036
standard deviation	0.506	0.506	0.422	0.351	0.280	2.756	16.099	1.133	0.253

In order to study the relationship between age and oxygen saturation, Pearson correlation coefficient is used to analyze.

General form of selection of correlation coefficients between various samples using Pearson algorithm:

Two sets of data:

$$X : \{X_1, X_2, \dots, X_n\} \quad Y : \{Y_1, Y_2, \dots, Y_n\}$$

Then the sample mean of the data:

$$E(X) = \frac{\sum_{i=1}^n X_i}{n}, E(Y) = \frac{\sum_{i=1}^n Y_i}{n}$$

Sample covariance:

$$Cov(X, Y) = \frac{\sum_{i=1}^n (X_i - E(X))(Y_i - E(Y))}{n}$$

Coefficient of sample Pearson:

$$\rho_{XY} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} = \frac{\sum_{i=1}^n (X_i - E(X)) \cdot (Y_i - E(Y))}{\sigma_X \sigma_Y n}$$

$\sigma_x$  These are x standard deviations and y standard deviations:  $\sigma_y$

$$\sigma_X = \sqrt{\frac{\sum_{i=1}^n (X_i - E(X))^2}{n}}, \sigma_Y = \sqrt{\frac{\sum_{i=1}^n (Y_i - E(Y))^2}{n}}$$

The scatter plot is drawn from Table I as follows:

and the average oxygen saturation of 36 experimental individuals were analyzed by multiple regression analysis, and the regression results were tested by hypothesis test, and the reasonable regression results satisfying the hypothesis test were obtained. The influencing factors of blood oxygen saturation were evaluated synthetically.

### 3.1.3 ASSUMPTIONS

Assume that the patient data are accurate.

Assume that all patients have no abnormal condition at the time of detection.

It is assumed that the individual singular data detected have no significant effect on the overall change of their data.

### 3.1.4 MODEL ESTABLISHMENT AND SOLUTION

Before using multiple regression, the mean value of oxygen saturation was calculated by descriptive statistics on its variance and body evaluation index, and its minimum, maximum, mean, median, skewness, kurtosis and standard deviation were calculated respectively. The following table is available:

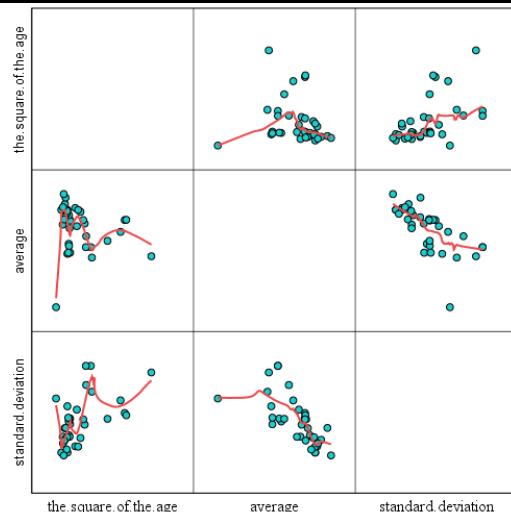


Figure 1: Data scatter plot of oxygen saturation

According to the scatter plot, the average value of age and oxygen saturation is negatively correlated to a certain extent, and the standard deviation of oxygen saturation is positively correlated. Based on this conjecture, the Pearson correlation coefficient is solved:

Sample mean of the data:

$$E(X) = \frac{\sum_{i=1}^n X_i}{n}, E(Y) = \frac{\sum_{i=1}^n Y_i}{n}$$

Sample covariance:

$$\text{Cov}(X, Y) = \frac{\sum_{i=1}^n (X_i - E(X))(Y_i - E(Y))}{n}$$

Coefficient of sample Pearson:

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{\sum_{i=1}^n (X_i - E(X)) \cdot (Y_i - E(Y))}{\sigma_X \sigma_Y n}$$

$$\sigma_X = \sqrt{\frac{\sum_{i=1}^n (X_i - E(X))^2}{n}}, \sigma_Y = \sqrt{\frac{\sum_{i=1}^n (Y_i - E(Y))^2}{n}}$$

The following table was calculated:

Table 2: The result of Table 1:

	square of the age	average	standard deviation
square of the age	1.0000	-0.1688	0.5382
average	-0.1688	1.0000	-0.6805
standard deviation	0.5382	-0.6805	1.0000

Table 3: Assumption test results

	square of the age	average	standard deviation
square of the age	1.0000	-0.1688	0.5382
average	-0.1688	1.0000	-0.6805
standard deviation	0.5382	-0.6805	1.0000

Under the condition of obtaining the above results, multiple linear regression is carried out.

The flow chart of multiple regression is as follows:

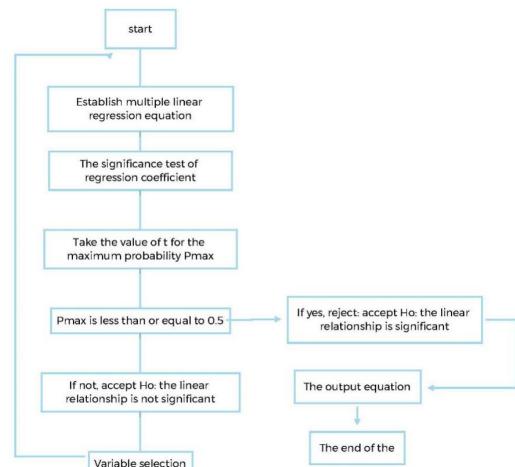


Figure 2: Flowchart of Multiple Regression

Normalization of raw data:

$$z_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \in [0, 1]$$

y Average and variance of oxygen saturation were x, body mass index, body health index, body comprehensive index I<sub>1</sub>x, I<sub>2</sub>x, I<sub>3</sub>, Assuming that the mean value of oxygen saturation is multivariate linear regression model with variance and body index, the regression model is constructed below.

Average:

$$y_1 = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon_i$$

Standard deviation:

$$y_2 = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon_i$$

ACADEMIC PUBLISHING HOUSE

If the original hypothesis is that the average oxygen saturation is independent of body indexes, the P values of body health index and body mass index are less than 0.05, which indicates that the two data reject the original hypothesis at 95% level. There is a correlation between oxygen saturation and the square number of body mass index and body health index.

The mean of oxygen saturation is related to body mass index ( $\beta_1=0.7120902$ ); hypothesis test: p value =0.000 less than 0.05

The average oxygen saturation was related to the physical health index ( $\beta_2=0.77531222$ ); hypothesis test: p value =0.000 less than 0.05

The correlation coefficient between oxygen saturation and body mass index is 0.71, which indicates that the body mass index does not increase by one unit, and oxygen saturation will increase by 0.71 percentage points. The correlation coefficient between oxygen saturation and body health index is 0.75, indicating that oxygen saturation will increase by 0.75 percentage points for each unit of square increase of body mass index.

In general, relative to the body mass index, the correlation coefficient between oxygen saturation and body health index is greater than oxygen saturation, which indicates that the body health index has a higher effect on the oxygen saturation of the sample. The oxygen saturation of the sample is mainly determined by its body health index. Overall regression results were better (prob=0.000<0.05) If the original hypothesis is that the average oxygen saturation is independent of the body indexes, the P values of the body health index, the body mass index and the body comprehensive index are less than 0.05, which indicates that the three data reject the original hypothesis at 95% level. That is, oxygen saturation is correlated with the three indexes.

The mean of oxygen saturation is related to the body mass index ( $\beta_1=0.15$ ); hypothesis test: the p value =0.000 is less than 0.05

The mean of oxygen saturation is related to the physical health index ( $\beta_2=0.12$ ); hypothesis test: the p value =0.000 is less than 0.05

The mean of oxygen saturation is related to the physical health index ( $\beta_3=-0.0004$ ); hypothesis test: p value =0.005 less than 0.05

The correlation coefficient between oxygen saturation and body mass index is 0.15, indicating that the standard deviation of oxygen saturation will increase by 0.15 percentage points for each unit of body mass index; the correlation coefficient between oxygen saturation and body health index is 0.12, indicating that the standard deviation of oxygen saturation will increase by 0.12 percentage points for each unit of square increase of body mass index; and the correlation coefficient of oxygen saturation with body composite index is 0.0004, indicating that for each unit of square increase of body composite index, the standard deviation of oxygen saturation will increase by 0.0004 percentage points;

In general, compared with the body health index and the body comprehensive index, the correlation coefficient between oxygen saturation standard deviation and body

mass index is larger, which indicates that the body mass index has a higher effect on the oxygen saturation standard deviation of the sample. The oxygen saturation standard deviation of the sample is mainly determined by its body mass index.

Table 2 shows that the correlation coefficient between the square of age and the average oxygen saturation is -0.1688, that is, the average oxygen saturation will decrease by 0.1688 percentage points for each unit of the square of age, and the correlation coefficient between the square of age and the standard deviation of oxygen saturation is 0.5832. According to the hypothesis test in Table 3, let the original hypothesis be that the average value of age and oxygen saturation is independent of the standard deviation of oxygen saturation. Because the p values of average and standard deviation are less than 0.05, the original hypothesis is rejected on 95% confidence interval.

Through the above analysis, we can know that age has a direct effect on the change of oxygen saturation of the sample. With age, the average oxygen saturation will decrease.

### 3.1.5 STRENGTH AND WEAKNESS

#### **Advantages:**

1. in this paper, the second principal component and the third principal component are squared to make the average value of oxygen saturation and the goodness of fit of variance reach the ideal value.
2. after obtaining the multivariate regression and body index functions, this paper again uses the Pearson correlation coefficient to solve the average value of age and oxygen saturation and the standard deviation. Finally, the conclusion similar to the regression function is drawn and complementary to make the result more objective.

#### **Disadvantages:**

Because principal component analysis itself is a screening process for data, although it can effectively reduce variables and achieve dimensionality reduction effect, the accuracy of data after principal component analysis is 81% of the total data. Further improvement is needed in this respect.

## 3.2 IMPROVED MODEL

### 3.2.1 MODEL IMPROVEMENT

Because of the OLS regression used in this paper, the following team will use different regression methods to fit it, such as Bayesian ridge regression, elastic network, support vector machine regression and so on, so that the fitting results are better

### 3.2.2 MODEL EXTENSION

This paper studies the specific correlation between the average value of oxygen saturation and the standard deviation of different body indexes, and provides important technical support for medical researchers. It can also be used to predict the impact of other physical indicators, and provide basic work support for medical and health workers.

## REFERENCES

- [1] Wang Qingrong. A Study and Application of Functional Principal Component Analysis and Functional Linear Regression Model [D]. Chongqing Industrial and Commercial University, 2020.
- [2] Jia Ninghua. A Study on Some Problems of Functional Linear Regression Model [D]. and Hefei University of Technology, 2012.
- [3] Ding Hui. A Study on the Functional Regression Model [D]. and East China Normal University, 2018.
- [4] Ding Hui, Xu Wenchao, Zhu Hanbing, Wang Guochang, Zhang Tao, Zhang Riquan. A Review of Functional Data Regression Analysis [J]. Application probability Statistics 34(06):630-654.
- [5] Liu Kesheng, Wang Siyang. Variable Selection [J]. in Independent Variable Regression Model with Function Type Journal of Beijing University of Aeronautics and Astronautics.
- [6] Wang Yan. A Study on Vector Autoregressive Association Prediction Model Based on Penalty Function [D]. and Central China normal University.
- [7] Wang Jiandu. on Variable Selection and Robust Parameter Design Based on Generalized Linear Model PhD Dissertation, 2011.
- [8] Wang Xiaojia. A Study on the Prediction Theory and Method Based on Data Analysis [D]. PhD Dissertation, 2012.

# Analysis of C4 Olefins Prepared by Ethanol Coupling

Ziao Xing<sup>1,2</sup> Jincan Wang<sup>1,3</sup> Wenhua Li<sup>1,3</sup>

<sup>1</sup>Association for Mathematics and Data Science, North China University of Science and Technology, Tangshan 063210, China;

<sup>2</sup>School of Materials Science and Engineering, North China University of Science and Technology, Tangshan 063210, China;

<sup>3</sup>School of Civil Engineering, North China University of Science and Technology, Tangshan 063210, China

**Abstract:** C4 olefins are widely used in medical and chemical fields. In this paper, the effects of 14 types of catalysts with CO/SiO<sub>2</sub> and HAP-ethanol concentration at different ratios and temperatures on the selectivity and ethanol conversion of C4 olefins by ethanol coupling were investigated. The trend of change was predicted by univariate analysis and curve fitting. It is concluded that 0.5wt%Co/SiO<sub>2</sub> and the ratio between Co/SiO<sub>2</sub> and HAP is 1:2, the C4 olefins selectivity and ethanol conversion are higher at ethanol concentration of 1.68 mL /min. The yield of C4 olefin was analyzed based on the data obtained from different catalyst combinations. The optimal catalyst combination under the temperature below 350°C and the optimal catalyst combination above 350°C were determined by taking 350°C as the boundary.

**Keywords:** C4 olefins selectivity; Univariate analysis interpolation

## 1.THE PROBLEM BACKGROUND

With the rapid development of industry, the lack of resources and environmental pollution, has gradually been paid attention to. Ethanol acts as a cleanerEnergy, can be obtained through straw, corn and other materials fermentation, raw material sources are very wide, the yield increased year by year, production. The cost decreases year by year. Ethanol can be used as raw material for C4 olefin, which is widely used in chemical products, medicine and so on Dabble in all areas. Selection of catalyst, reaction temperature and reaction time in the preparation of olefin by ethanol catalytic coupling The length of the reaction is greatly related to the loss of raw materials and the reaction rate [1]. Therefore, explore the difference between temperature and no The optimal conditions under the combination of catalysts and different time are particularly important.

## 2. DATA PREPROCESSING AND ANALYSIS

Code name	A1	A2	A3	A4	A5
	200mg 1wt%Co/ SiO <sub>2</sub> -200mg HAP-ethanol concentration 1.68 mL /min	200mg 2wt%Co/ SiO <sub>2</sub> -200mg HAP-ethanol concentration 1.68 mL /min	200mg 1wt%Co/ SiO <sub>2</sub> -200mg HAP-ethanol concentration 0.9 mL /min	200mg 0.5wt%Co/ SiO <sub>2</sub> -200mg HAP-ethanol concentration 1.68 mL /min	200mg 2wt%Co/ SiO <sub>2</sub> -200mg HAP-ethanol concentration 0.3 mL /min
Code name	A6	A7	A8	A9	A10
species	200mg 5wt%Co/ SiO <sub>2</sub> -200mg HAP-ethanol concentration 1.68 mL /min	50mg 1wt%Co/ SiO <sub>2</sub> -50mg HAP-ethanol concentration 0.3 mL /min	50mg 1wt%Co/ SiO <sub>2</sub> -50mg HAP-ethanol concentration 0.9 mL /min	50mg 1wt%Co/ SiO <sub>2</sub> -50mg HAP-ethanol concentration 2.1 mL /min	50mg 5wt%Co/ SiO <sub>2</sub> -50mg HAP-ethanol concentration 2.1 mL /min
Code name	A11	A12	A13	A14	
species	50mg 1wt%Co/SiO <sub>2</sub> 90mg quartz sand-ethanol concentration without HAP 1.68ml/min	50mg 1wt%Co/ SiO <sub>2</sub> -50mg HAP-ethanol concentration 1.68 mL /min	67mg 1wt%Co/ SiO <sub>2</sub> -33mg HAP-ethanol concentration 1.68 mL /min	33mg 1wt%Co/ SiO <sub>2</sub> -67mg HAP-ethanol concentration 1.68 mL /min	

First, set codes for various catalysts as shown in Table1 belowTable1.Category number

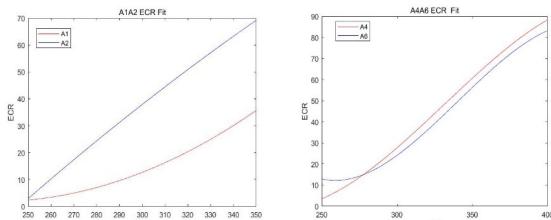


Figure1. Ratio of Co/SiO<sub>2</sub> weight to ethanol and C4 olefin conversion

In the experiment, the method of controlling single variable was used to divide the catalysts into four categories. Firstly, A1, A2, A4 and A6 were divided into

one category to explore the influence of the weight ratio of Co/SiO<sub>2</sub> on the conversion of ethanol and C4 olefin. As shown in Figure1.

According to the observation of Figure1, it can be clearly seen that under any conditions of weight ratio of Co/SiO<sub>2</sub>, the ethanol conversion rate has an obvious trend of increase with the increase of temperature. However, under the catalytic condition of A1, the growth rate of ethanol conversion rate is relatively slow compared with the other three groups, only reaching 35% at 350°C, far lower than other groups. At the same time, taking 350°C as the boundary, we can see that at 350°C, the ethanol conversion rate of A2 is the maximum, and the curve has a trend of slowing down. Therefore, we can infer that under 350°C, the optimal catalyst combination of A2:

Co/SiO<sub>2</sub> ratio is 2:100, i.e. 2wt%Co/SiO<sub>2</sub>. However, according to the curve fitting, it can be observed that the ethanol conversion under the condition of A2 catalyst has a trend of slowing down. Under the condition of A4 and A6 catalyst, the ethanol conversion rate under the conditions of A4 and A6 catalyst continues to increase to 400°C at a rapid rate. At the same time, the ethanol conversion rate under the condition of A4 combined catalyst is greater than that under the condition of A6, and the curve of A6 shows a trend of slowing down. Therefore, at 400°C, the optimal catalyst ratio of Co/SiO<sub>2</sub> is 1:2000, that is, 0.5wt%Co/SiO<sub>2</sub>, to ensure higher ethanol conversion. This is consistent with the references we consulted.[2] The selectivity growth rate of A1A2A4A6 for C4 olefins was plotted as Figure2.

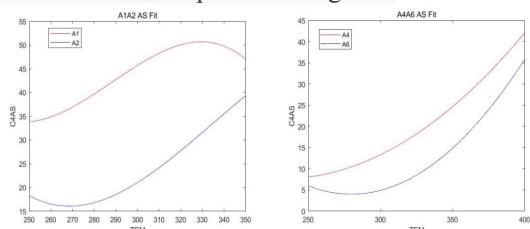


Figure2. Effect of A1A2A4A6 on the growth rate of C4 olefins selectivity

As can be seen from Figure2, with the increase of temperature, the selectivity of C4 olefin increases under different catalyst combinations, but the range is small. Meanwhile, at 320°C, A1 catalyst has a downward trend, which reaches up to about 50%. A4, A6 two types of catalyst temperature makes C4 olefin selectivity of gradually rise, if you can choose when low temperature A1 type of catalyst, to improve the C4 olefin selectivity, indirect improve C4 hydrocarbon output, but if the temperature under the condition of high slope can choose C4 olefin growth larger A6 catalyst make C4 selective enhancement, Thus increasing the output of C4 olefin.

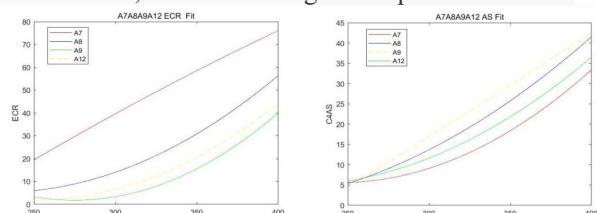


Figure3. Influence of HAP and ethanol concentration on ethanol conversion and C4 olefins selectivity

According to Figure3, we conducted variable analysis on HAP and ethanol concentration, and classified A7, A8, A9 and A12 into one category. According to Figure3, it can be seen that under the conditions of these four catalysts, ethanol conversion rate and C4 olefins selectivity increased with the increase of temperature, and the increase range was large. But compared with other catalysts combination in A9 catalyst under the condition of combination of ethanol conversion rate of growth is relatively flat, but in the process of temperature rise, the C4 olefin selectivity is always higher than other catalysts, so we think: under the condition of high temperature, A9 catalyst composition also has a good C4 hydrocarbon catalytic ability of output.

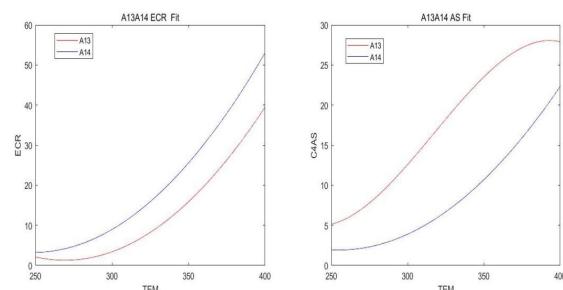


Figure4. Different ratios between Co/SiO<sub>2</sub> and HAP affect the results

As can be seen from Figure4, A13 and A14 are finally grouped into one group to explore the influence of different ratios between Co/SiO<sub>2</sub> and HAP on the results. According to Figure4, with the increase of temperature, the conversion rate of ethanol and the selectivity of C4 olefin are both on the whole rising. At the same time, according to the fitting curve, we can observe that the influence curve of A13 and A14 catalyst on ethanol conversion is similar. However, the combination of A13 and A14 catalysts has different selectivity effects on C4 olefin. Within 400°C, the selectivity of C4 olefin has been increasing under the condition of A14 catalyst combination, and the growth rate is also increasing, while the combination of A13 catalyst has passed the maximum point of C4 olefin selectivity between 350°C and 400°C. The trend gradually flattens, and even the C4 olefins selectivity decreases, which is not conducive to the further production of C4 olefins. Therefore, we believe that with the increase of temperature, the A13 catalyst combination, i.e. the ratio between Co/SiO<sub>2</sub> and HAP, is 67:33, which is more suitable for the catalytic coupling of ethanol to prepare C4 olefin.

### 3.YIELD ANALYSIS OF C4 OLEFIN

C4 olefin yield = ethanol conversion \*C4 olefin selectivity

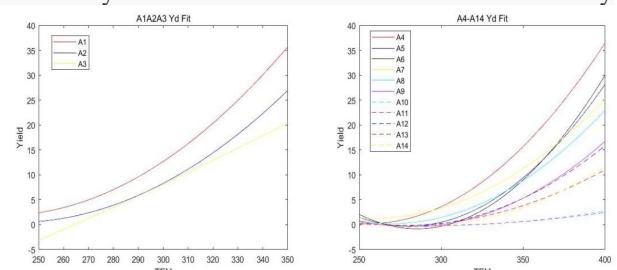


Figure5. Influence of catalyst combination on C4 olefin yield

As the temperature difference is different and the experimental description is scientific, we conduct curve fitting for the data according to different temperature differences. According to the observation and analysis in Figure5, it can be seen that at 350°C, the yield curves of C4 olefin of A1 catalyst combination and A2 catalyst combination do not show a trend of slowing down or decreasing. Therefore, we speculate that the yield of C4 olefin will continue to increase with the temperature continuing to rise. At the same time, it can be seen that after 350°C, the yield of C4 olefin in each catalyst combination still increases with the increase of temperature. Finally, at 400°C, the yield of C4 olefin in

catalyst combination A4: Under the condition of 200Mg0.5 wt%Co/ SiO<sub>2</sub>-200MgHAP-ethanol concentration of 1.68 mL /min, the selectivity of C4 olefins was significantly higher than that of other catalyst combinations, and with the increase of temperature, the yield span of C4 olefins was the largest under the condition of the A4 catalyst combination, and it can be inferred that within the temperature range of 250°C to 400°C, The A4 catalyst combination is the most sensitive to temperature. Meanwhile, based on Figure1 and Figure2, the optimal temperature and catalyst combination can be determined as 400°C, 0.5wt%Co/ SiO<sub>2</sub>-200MgHAP-ethanol concentration: 1.68mL /min.

#### 4.CONCLUSION

In this paper, through the influence of 14 kinds of catalysts on the conversion of ethanol and the yield of C4 olefin, it is found that the selectivity of C4 olefin is greater and the conversion of ethanol is higher under the condition of 0.5wt%Co/SiO<sub>2</sub>, the ratio between Co/SiO<sub>2</sub> and HAP is 1:2, and the ethanol concentration is 1.68 mL /min. We can also determine the optimal temperature and catalyst combination for C4 olefins yield: 400°C, 0.5wt%Co/ SiO<sub>2</sub>-200mgHAP-ethanol concentration: 1.68 mL /min.

#### 5.OUTLOOK

Based on different kinds of catalyst for C4 olefin selectivity, as well as the effect of ethanol conversion of

exploration, we found that the existence of a large number of side effects, we should be on how to improve the ethanol conversion rate at the same time, reduce the side effects, so as to increase the production rate of C4 olefin, this will be our faces a very challenging problem. At the same time, through the above exploration, we found that C4 olefin can be prepared at a lower temperature, so we can apply different catalysts to chemical production according to their catalytic performance and the optimal catalytic temperature. Through the exploration and mining of this experimental data, it opens up new ideas for the utilization of ethanol and the development of clean energy. With the development and modification of more catalyst combinations in the future, we believe that China will gradually achieve the expected goal of energy conservation and emission reduction.

#### REFERENCES

- [1] Chen G, Kirschning A. First Preparation of Spacer-Linked Cyclic Neooligoaminodeoxysaccharides[J]. Chemistry—A European Journal, 2002, 8(12): 2717-2729.
- [2] Lv Shaopei. Synthesis of butanol and C4 olefin by ethanol coupling [D]. Dalian University of Technology, 2018.

# Optimization Analysis of C4 Olefin Preparation by Ethanol Coupling

Yiming Xu<sup>1, 2</sup>, Haoyu Zhou<sup>1, 2</sup>, Tengda Zhang<sup>1, 2</sup>

<sup>1</sup>Mathematical Modeling Innovation Lab, North China University of Science and Technology, Tangshan 063210, China;

<sup>2</sup>Architectural Engineering Institute Innovation Education North China University of Science and Technology, Tangshan 063210, China

**Abstract:** Aiming at the problem of coupling ethanol to prepare C4 olefins, the relationship and influence of ethanol conversion, C4 olefin selectivity and temperature were studied based on fitting and classification. Then, the control variable method is used and a regression analysis model is established to determine the optimal catalyst combination and the temperature to maximize the C4 olefin yield. Finally, the experimental process is designed and improved.

**Key words:** SPSS Data Processing; Multiple Regression Analysis; Control Variable Method; Random Forest Prediction; Goal Programming

## 1. PROBLEM RESTATEMENT

### 1.1 PROBLEM BACKGROUND

The processing and deep processing of C4 olefins have always been paid close attention by the industry [1]. The preparation of C4 olefins is of great significance and value because it is widely used in chemical products, medicine and pharmacy, driving the development of oil refining and chemical industry chain. Catalytic coupling of ethanol to C4 olefins is an important method. The process conditions have high requirements and standards. Catalyst combination and temperature are the main factors affecting the selectivity and yield of C4 olefins. Therefore, the key to solve this problem is to study and design the catalyst combination and the preparation efficiency of C4 olefins at a specific temperature.

The production of C4 olefins from ethanol condensation. The reaction mechanism is different on different catalysts. When the catalyst is mixed metal oxide, the mixed metal oxide has appropriate acid and base sites, catalyzes ethanol dehydration at the acid sites, and the basic sites promote alcohol dehydrogenation and the coupling of intermediates. The surface of SiO<sub>2</sub> HAP catalyst is supported with Co metal with dehydrogenation activity. By adjusting the loading amount of Co, the acid and alkalinity of the catalyst surface can be adjusted, so that SiO<sub>2</sub> HAP catalyst has acid and alkali activities at the same time. In the experiment, the catalyst first catalyzes ethanol to produce acetaldehyde and acetone intermediates, and acetaldehyde and acetone produce C4 olefins such as isobutene and butadiene through further coupled dehydration.

### 1.2 PROBLEM RESTATEMENT

Question 1: the performance data sheet has given the values of ethanol conversion and C4 olefin selectivity at different temperatures corresponding to different loading methods and catalyst combinations. Based on the data in

Annex 1 above, the relationship between ethanol conversion, C4 olefin selectivity and temperature is studied respectively. The test results of reaction products corresponding to different times under specific catalyst combination at 350 °C in Annex 2 are studied.

Problem 2: analyze the effects of different catalyst combinations and temperatures on ethanol conversion and C4 olefin selectivity.

Problem 3: in order to make the higher the C4 olefin yield under the same experimental conditions, determine the appropriate catalyst combination and temperature conditions; Under the condition that the temperature is lower than 350 °C, the appropriate catalyst combination and temperature conditions are determined in order to make the C4 olefin yield higher and better.

Question 4: five new experiments are designed to explore the preparation of C4 olefins by ethanol coupling.

For problem 1: Based on the data in Annex I, this topic uses Excel to visualize the data, roughly judge the trend through the broken line diagram, and then establish a data fitting model to study the relationship between ethanol conversion, C4 olefin selectivity and temperature for each catalytic combination. It is concluded that for different catalyst combinations, the graphic trend presented by the functional formula can be treated in groups, that is, with the change of temperature, the change mode of ethanol conversion is three, linear, S-type and progressive type; The selectivity of C4 olefins changes in two ways, S-type and progressive type. Then, a linear regression model is established to fit and analyze the results of Annex II. The relationship between ethanol conversion and time is  $y_1 = 42.671 - 0.053t$ , The selectivity of C4 olefins with time is.  $y_2 = 16.53 - 0.02t$ .

Aiming at problem 2: the influence of the composition and temperature of each catalyst combination is obtained by using the control variable method. The conclusions are as follows: 1. The co loading increased from 0.5wt% to 5wt%, the conversion of ethanol decreased first and then increased, and the selectivity of C4 olefins increased first and then decreased. In general, the yield of C4 olefins increased first and then decreased, and reached the highest when the loading was 2wt%. 2. When the ratio of CO / SiO<sub>2</sub> to HAP is 200mg: 200mg, the experimental effect is the best. Reducing the dose will reduce the effect. With the increase of HAP ratio (the ratio increases from 1 / 2 to 2), the selectivity of C4 olefins gradually decreases, and ethanol is more converted to alcohols. 3. The mixing method has little effect. 4. The ethanol conversion decreased with the increase of ethanol flow rate, and the

C4 olefin yield was not greatly affected by ethanol flow rate. The effect is better when the flow rate is 0.9ml/min and 2.1ml/min. 5. HAP is better than quartz sand for catalyst support.

For problem 3: calculate the C4 olefin yield, select the top ten groups with the largest yield, and combined with the analysis of the second question, it is concluded that the best catalyst group is 200mg 1wt% Co / SiO<sub>2</sub> - 200mg HAP ethanol concentration of 0.9ml/min. The optimum temperature is 400 °C. If the temperature is lower than 350 °C, arrange and combine the catalyst combination variables to obtain 36 groups of experiments. The yield of C4 olefins in 36 groups of experiments is predicted by using the optimized random forest model. Combined with the original data, it is concluded that the best catalyst group is 200mg 2wt% Co / SiO<sub>2</sub> - 200mg HAP ethanol concentration of 1.68ml/min. The optimum temperature is

Order	The Proportion of Co/SiO <sub>2</sub> And HAP	Catalyst support
1	200mg:200mg	HAP
2	200mg:200mg	HAP
3	200mg:200mg	HAP
4	200mg:200mg	HAP
5	200mg:200mg	HAP

## 2.PROBLEM ANALYSIS

To solve problem 1: in the process of ethanol coupling to prepare C4 olefins, ethanol accelerates the reaction under the action of catalyst to produce a variety of chemicals, including C4 olefins. Annex I has given 21 catalyst combinations and their corresponding proportions of chemicals at different temperatures. Based on this, the relationship between ethanol conversion and temperature, C4 olefin selectivity and temperature was found through Excel data analysis and mapping. Then, a linear regression model was established by SPSS to study the results of 350 °C specific catalyst combination in Annex II. The relationship between ethanol conversion and C4 olefin selectivity at this temperature and different reaction time was found.[1] For problem 2: each catalyst combination has specific co loading, Co / SiO<sub>2</sub> and HAP loading ratio, HAP and ethanol concentration. And with the change of temperature, the corresponding values of each index also change. Therefore, using the control variable method, 21 combinations were divided into several groups for comparative analysis to study their effects on ethanol conversion and C4 olefin selectivity.

For question 3: Based on the results of the second question, calculate the C4 olefin yield and take out the top ten with the highest yield. If the temperature is lower than 350 °C, several groups of data of Co loading, ethanol flow rate and temperature are arranged and combined, and the optimized random forest model is used to predict the yield of C4 olefins respectively, taking the top ten groups with the largest value.

Aiming at problem 4: Based on the above problems, it is analyzed that temperature and catalyst combination are important factors affecting the experiment. Five groups of experiments were added to predict C4 yield by studying

350 °C. At the same time, a nonlinear regression model is established to study the relationship between CO load, ethanol flow rate and temperature and yield. Then, the objective programming model is used to solve it. The results are approximately consistent with the above analysis.

Aiming at problem 4: Based on the above problems, it is analyzed that temperature and catalyst combination are important factors affecting the experiment. Five groups of experiments were added to predict C4 yield by studying random forest algorithm. Among them, two experimental groups with a temperature higher than 450 °C are set to explore whether excessive temperature will reduce the catalyst activity. Three experiments are designed near the best catalyst group to explore whether there is a requires dehydrogenation, dehydration and better catalyst combination. The five experimental designs are as follows:

Order	The Proportion of Co/SiO <sub>2</sub> And HAP	Catalyst support	The Load of Co wt%	Ethanol flow rate ml/min	Temperature
1	200mg:200mg	HAP	1	0.9	475
2	200mg:200mg	HAP	0.5	1.68	450
3	200mg:200mg	HAP	1	0.3	400
4	200mg:200mg	HAP	1	2.1	400
5	200mg:200mg	HAP	1	1.68	400

random forest algorithm. Among them, two groups were set at a temperature higher than 450 °C, and three groups of experiments were designed near the best catalyst group to explore whether there is a better catalyst combination.

## 2. MODEL ASSUMPTIONS

It is assumed that all experimental data are completed without human interference.

It is assumed that the instruments of all experiments are exactly the same.

It is assumed that the composition and dosage of the catalyst combination are accurate.

It is assumed that the time record in Annex 2 is accurate and does not include human reaction time

## 3. SYMBOL DESCRIPTION

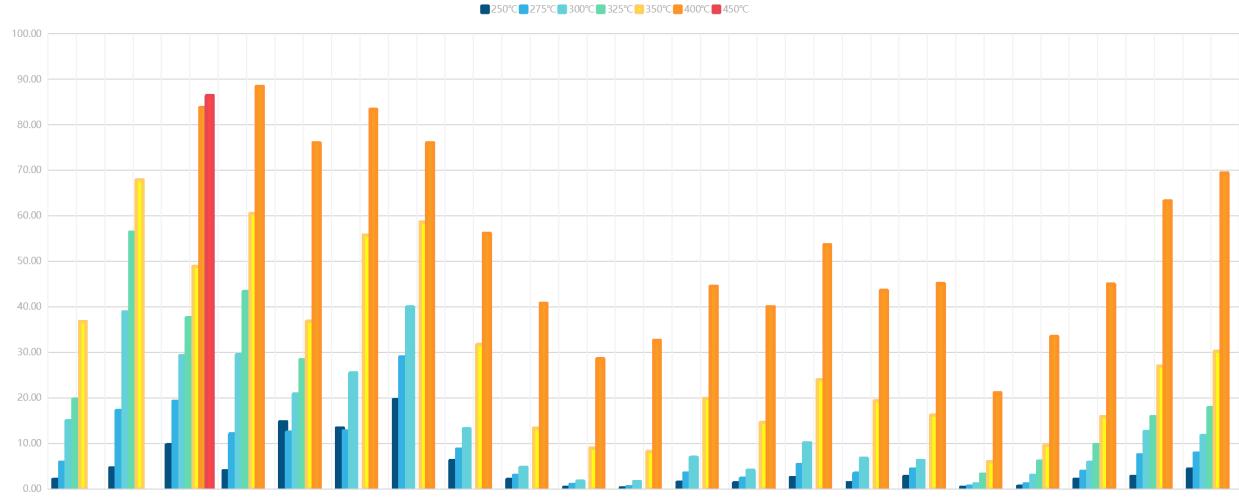
The main mathematical operation symbols and their meanings are shown below:

Symbol	meaning
$y_1$	Ethanol conversion
$y_2$	The Selectivity of C4 olefin
$t$	temperature
$x_1$	The Quality of Co
$x_2$	The Load of Co
$x_3$	Ethanol flow rate

## 4.ESTABLISHMENT AND SOLUTION OF MODEL

### 5.1QUESTION 1

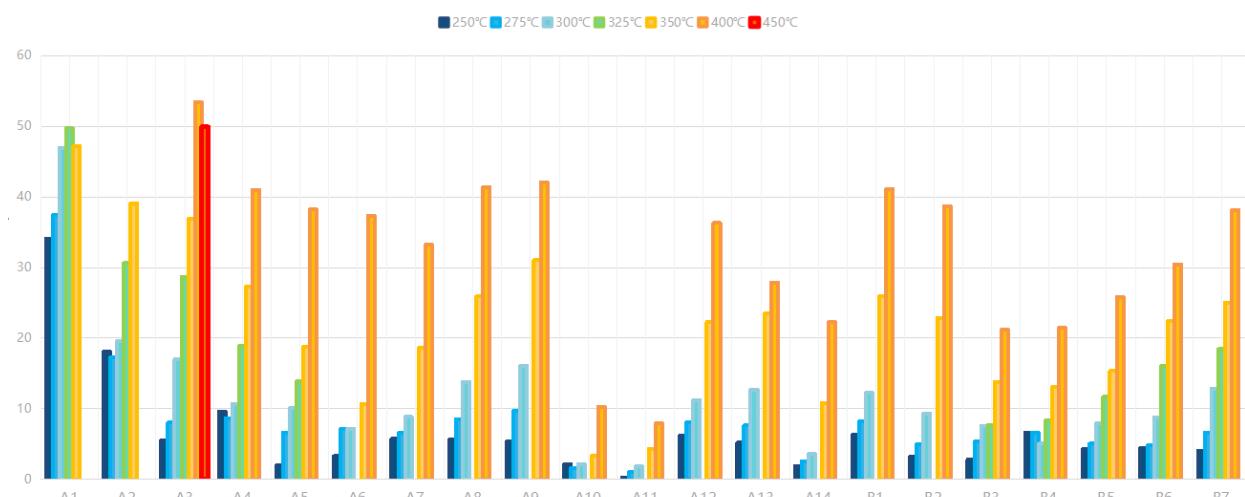
Relationship between ethanol conversion and temperature In the process of preparing C4 olefins by ethanol coupling, other products other than C4 olefins are produced, such as ethylene, acetaldehyde, fatty alcohol with carbon number of 4-12 and other chemicals. The ethanol conversion of different catalyst combinations is different at a specific temperature. As shown in the following chart.



As shown in the figure above, the ethanol conversion of different catalyst combinations increases with the increase of temperature. When the temperature increased to 450 °C, the combination A3 accelerated the reaction, and the ethanol conversion was 86.4%, which was higher than that at 400 °C.

#### Temperature dependence of C4 olefin selectivity

The longitudinal comparison is shown in the figure below. With the increase of temperature, the greater the selectivity of C4 olefins under almost every catalyst combination. The catalyst combination A3 has the best experimental effect at 400 °C, and the temperature further increases. When it reaches 450 °C, the selectivity of C4 olefins decreases.



Considering the relationship between ethanol conversion, C4 olefin selectivity and temperature, it is found that at 400 °C, under the action of catalyst combination A3, i.e. 200mg-1wt% Co / SiO<sub>2</sub>-200mg HAP ethanol concentration of 0.9ml/min, ethanol coupling has high conversion and high selectivity, which is the best process condition under qualitative analysis.[2]

#### quantitative analysis

First, observe the approximate relationship between ethanol conversion and C4 olefin selectivity of each group of catalysts and temperature with the data perspective. It is found that the relationship between ethanol conversion and temperature can be divided into three categories:

Type s, including A1, A2, A4, A7;

Type s, including A3;

Progressive type, including A5, A6, a8 ~ A14 and B1 ~ B7.

Then, A4, A3, A6 and B3 are selected as representatives for fitting analysis according to the broken line diagram to

determine the functional relationship.

A4	$f(t) = 0.58t - 144.6$
A3	$f(t) = 46.66 - 15.65 \cos(0.0199t) + 36.35 \sin(0.0199t) + 3.296 \cos(0.0398t) - 9.75 \sin(0.0398t)$
A6	$f(t) = 49.8 - 3.73 \cos(0.0187t) + 37.17(0.0187t)$
B3	$f(t) = 424.7 e^{-\frac{(t-664.6)}{152.7}}$

Through comparison, it is found that the growth rate of conversion increases rapidly after 350 °C, but the value of group B is generally small.

The relationship between C4 olefin selectivity and temperature can be divided into two categories:

Type s, including A1 and A3;

Progressive type: including A2, A5 ~ A14 and B1 ~ B7.

Similarly, A1, A3, A2 and A5 are selected as representatives for fitting analysis to determine the functional relationship.

A1	$f(t) = 42.16 + 6.189 \cos(0.04t) + 5.48 \sin(0.04t)$
A3	$f(t) = 29.57 + 7.831 \cos(0.018t) + 23.34 \sin(0.018t)$

A2	$f(t) = -1.795 \times 10^{-6}t^4 + 0.00212t^3 - 0.9342t^2 + 181.2t - 1.307 \times 10^4$
A5	$f(t) = 194e^{(\frac{t-668.7}{210.6})^2}$

However, the value of A1 is generally high, the maximum at 325 degrees, and then decreases slowly; A3 has a fast growth rate and a large growth range, which is the largest at 400 degrees, and then decreases rapidly; A2 increases progressively and rapidly after 300 °C; A5 increased progressively and rapidly after 350 °C.

Finally, a linear regression model was established with time as the independent variable and ethanol conversion as the dependent variable.

The regression model of ethanol conversion and time is as follows:

$$y_1 = 42.671 - 0.053t$$

Since the goodness of fit  $R^2 = 0.933$ , the fitting effect of the regression model is good. That is, the explanation degree of the change of ethanol conversion over time is 93.35%. The result of analysis of variance was  $p < 0.05$ , indicating that the model was statistically significant.

The analysis of the change of C4 olefin yield with time is the same. There are no outliers in the data and meet the conditions of simple linear regression.[3]

The regression model of C4 olefin yield and time is as follows:

$$y_2 = 16.53 - 0.02t$$

Since the goodness of fit  $R^2 = 0.857$ , the fitting effect of the regression model is good. That is, the explanation degree of the change of ethanol conversion over time is 85.7%. The result of analysis of variance was  $p = 0.003 < 0.05$ , indicating that the model was statistically significant.

## 5.2 QUESTION 2

By analyzing the experimental groups, we classified the experimental catalyst combinations, and discussed the effects of different catalyst combinations and temperatures on ethanol conversion and C4 olefin selectivity from the conditions of Co loading,  $\text{SiO}_2 / \text{HAP}$  ratio, mixing mode, reaction temperature and ethanol flow rate.

The Co loading increased from 0.5wt% to 5wt%, the conversion of ethanol decreased first and then increased, and the selectivity of C4 olefins increased first and then decreased. In general, the yield of C4 olefins increased first and then decreased. When the loading was 2wt%, it reached the highest, 26.54%. The co loading amount is 5wt%, which is worse than other effects.

When the catalyst ratio is 1:1, with the increase of HAP content (the dose increases from 50mg to 200mg), the content of basic sites in the catalyst gradually increases, which promotes the coupling conversion of acetaldehyde to olefinic aldehyde. At this time, the catalyst surface has a suitable acid that can promote the conversion of olefinic aldehyde to C4 olefin, so the selectivity of C4 olefin is the highest, 47.21% (200mg: 200mg). When the proportion of HAP is increased, the surface acidity dimension of the catalyst decreases, and the selective hydrogenation of olefinic aldehydes is further coupled to produce alcohols (fatty alcohols with carbon number of 4-12).

When the ratio of Co /  $\text{SiO}_2$  to HAP is 200mg:200mg, the experimental effect is the best, and reducing the dose will

reduce the effect. With the increase of HAP ratio (from 1 / 2 to 2), the selectivity of C4 olefins decreases gradually, and ethanol is more converted to alcohols.[4]

Under the same reaction conditions, there is little difference in ethanol conversion and C4 olefin selectivity between the two catalysts, indicating that the loading mode has little effect on the performance of the catalyst. We can conclude that the ethanol conversion decreases with the increase of ethanol flow rate, and the C4 olefin yield is not greatly affected by ethanol flow rate. The effect is better when the flow rate is 0.9ml/min and 2.1ml/min. When the flow rate is 0.3ml/min and 2.1ml/min, higher temperature will make the effect worse. Under the same reaction conditions, the effect of using HAP as carrier is better than that of using quartz sand as catalyst carrier, whether for ethanol conversion or C4 olefin selectivity.

## 5.3 QUESTION 3

The catalyst combination was analyzed separately and the existing experimental data were studied. It was found that the three variables of Co loading, reaction temperature and ethanol flow rate showed a quadratic nonlinear functional relationship with C4 olefin yield, that is, when other conditions were certain, it showed a trend of first increasing and then decreasing. After comparison, it can be reasonably inferred that the combination of Co /  $\text{SiO}_2$  and HAP with a charge ratio of 1 and a mass of 200mg is the best. Therefore, the model can be established only by studying the specific functional relationship between Co loading, reaction temperature and ethanol flow rate and C4 olefin yield.

Temperature is closely related to yield, ethanol flow rate is negatively related to the first trip, and the correlation of Co load is small. The corresponding regression analysis model can be established.

Based on the data trend, we select the quadratic nonlinear relationship to establish the regression model, and set the three variables of Co load, reaction temperature and ethanol flow rate as X2, t and X3, from which we get:

$$y = 30.06 + 2.52x_2 + 29.34t - 0.4x_3 - 0.17x_2^2 - 4.15t^2 + 0.001x_3^2 - 0.64x_2t - 0.001x_2x_3 - 0.068tx_3$$

Establish goal planning model:

$$\max y = f(x)$$

$$\begin{cases} 0.5 \leq x_2 \leq 5 \\ s.t. \quad 250 \leq t \leq 450 \\ 0.3 \leq x_3 \leq 2.1 \end{cases}$$

The calculation results are approximately as follows: Co load  $x_2 = 200\text{mg}$ , temperature  $T = 400$  °C, ethanol flow rate  $x_3 = 0.9\text{ml/min}$ .

## 5.4 QUESTION 4

Five groups of experiments were added

As the selectivity of C4 olefins in groups A1 and A3 decreases when the temperature is too high, but the overall yield of C4 olefins. From the original experimental data, the higher the temperature, the greater the yield of C4 olefins. We consider whether the temperature will affect the catalyst activity and whether the temperature higher than 450 °C will affect the yield of C4 olefins.

In theory, four groups of experiments were designed, and the prediction results using the optimized random forest model were as follows: the predicted temperature of group A1 and A2 was 400 °C, the predicted temperature of group A3 was 475 °C, and the predicted temperature of group A4

was 450 °C (the ratio of Co / SiO<sub>2</sub> to HAP was 200mg:200mg, and the catalyst carrier was HAP). The results of catalyst combination, temperature and predicted C4 olefin yield are as follows:

Order	The Proportion of Co/SiO <sub>2</sub> And HAP	Catalyst support	The Load of Co wt%	Ethanol flow rate ml/min	Temperature °C
1	200mg:200mg	HAP	1	0.9	475
2	200mg:200mg	HAP	0.5	1.68	450
3	200mg:200mg	HAP	1	0.3	400
4	200mg:200mg	HAP	1	2.1	400
5	200mg:200mg	HAP	1	1.68	400

## 5. ADVANTAGES AND DISADVANTAGES OF THE MODEL

### 6.1 ADVANTAGE

The linear regression model in question 1 considers the assumptions of linearity, independence and residual normality, and the results are accurate.

This paper visualizes the data, which is concise and practical.[5]

The system considered by the model is comprehensive and has strong practicability.

### 6.2 SHORTCOMING

1. This paper ignores the influence of other reaction products such as ethylene, acetaldehyde, methylbenzaldehyde and methylbenzyl alcohol, which reduces the accuracy of the model.

2. In the analysis of control variable method, the model is more complex because all components of catalyst combination are considered.

## 4. GENERALIZATION OF MODEL

In this paper, the effects of catalyst combination and temperature on the yield of C4 olefins were studied by controlling variable method. The analytical method is widely used and can be extended to chemical research, drug preparation and so on. A linear regression model was also established to study the relationship between ethanol conversion, C4 olefin selectivity and temperature. The model can be extended to study the relationship between dependent variables and independent variables, which is intuitive, concise and practical.

## REFERENCE

[1] Hao Zhuang, Guoping Huang, Yashuai Jiang, Guan Sun, Xueliang Bai, Jingnan Li. Adhesive Reliability Study at the Encapsulant/Backsheet Interface of Solar Modules for Various Ethylene-Vinyl Acetate Copolymer White Films[A]. Advanced Science and Industry Research Center.Proceedings of 2020 International Conference on

New Energy, Power and Environmental Engineering(NEPEE2020)[C].Advanced Science and Industry Research Center:Science and Engineering Research Center, 2020:5.

[2] Ya Nan Xie, Zhi Meng, Fu Li. Purification Process Design of Vinyl Acetate Based on Aspen Plus[A]. AEIC Academic Exchange Information Center(China).Proceedings of 2020 International Conference on Green Chemical and Environmental Science(ICGCES 2020)[C].AEIC Academic Exchange Information Center(China):International Conference on Humanities and Social Science Research, 2020:8.

[3] Xiaolan Liu, Yong Zheng. Experimental Study on Inhibition of Hydrate Formation by Compound Solution of PVP and Ethylene Glycol Ether[A]. Advanced Science and Industry Research Center.Proceedings of 2019 5th International Conference on Green Materials and Environmental Engineering(GMEE 2019)[C].Advanced Science and Industry Research Center:Science and Engineering Research Center, 2019:6.

[4] Fernando, N., Kathrotia, T., Fradet, Q., Braun-Unkhoff, M., Riedel, U.. The Oxidation of Ethylene Glycol-A Pyrolysis Bio Oil Surrogate[A]. ETA-Florence Renewable Energies.Proceedings of the 26th European Biomass Conference and Exhibition[C].ETA-Florence Renewable Energies:ETA-Florence Renewable Energies, 2018:1.

[5] Moura Vilela, C.F., Aranda Almansa, G., Vreugdenhil, B.J., Bos, A., Kuo, Y.T.. Co-Production Schemes in the Bio-SNG Process: BTX Production and Harvesting[A]. ETA-Florence Renewable Energies.Proceedings of the 26th European Biomass Conference and Exhibition[C].ETA-Florence Renewable Energies:ETA-Florence Renewable Energies, 2018:1.

# Research And Development of Suitable Aging Platform Based on Middle-Aged And Elderly Online Shopping Market Research

*Haowei Yan<sup>1,2</sup>, Xiaoru Huo<sup>2</sup>, Jiajia Liu<sup>1,3</sup>*

<sup>1</sup>*Laboratory of Engineering Calculation and Simulation Innovation, North China University of Technology, Tangshan 063210, China;*

<sup>2</sup>*School of Architectural Engineering, North China University of Technology, Tangshan 063210, China;*

<sup>3</sup>*College of Science, North China University of Technology, Tangshan 063210, China*

**Abstract:** With the penetration of the Internet, the middle-aged and elderly online shopping market outside the 'digital economy spotlight' has great potential. Based on this situation, this paper first uses stratified and two-stage sampling to extract the middle-aged and elderly population in Tangshan, and investigates online shopping through interview-based survey method. Secondly, using correspondence analysis, ordered multi-class Logistic regression and correlation analysis to understand the status of middle-aged and elderly online shopping. Through K-means++, the middle-aged and elderly groups are subdivided and the portraits of users are depicted. Finally, based on the survey results, build a concise page, with voice input and playback, double audit, flow pool intelligent recommendation function of the transfer Taobao, Pinduoduo and other platforms of the middle-aged and elderly shopping APP, for the middle-aged and elderly into the online shopping market and economic development to provide direction.

**keyword:** Correspondence Analysis; Ordered Multi-Classification Logistic Regression; K-Means++; Relevant Analysis; E-Commerce Platform

## 1. INTRODUCTION

National statistics show that China has become the country with the largest elderly population in the world. At the same time, China's aging process is developing rapidly with an annual increase of 10 million people. It is expected that the proportion of the elderly population aged 65 and above will reach 14% in 2020, and the whole will enter a deep aging society. The aging problem has brought pressure to China's economy and also brought opportunities. The expansion of the proportion of the population and the improvement of living standards make today's elderly groups have the consumption potential that cannot be ignored. The development and application of the Internet model in daily life provides more choices for their consumption. More types of goods and more convenient logistics under the e-commerce model also provide new exports for the exploration of their consumption potential.

In recent years, the network share of middle-aged and elderly people has gradually increased. According to Tencent's official data, as of September 2018, there were as many as 61 million WeChat users aged between 55 and 70. In various forms of network, pictures, audio and video

are obviously more favored by the middle-aged and elderly people. Correspondingly, new forms of traffic such as live broadcast and short video carrying commodity economy will also become a new trend for middle-aged and elderly people to shop. And according to the 'Focus on the Silver Economy-2019 Network Consumption Trend Report for Middle-aged and Elderly People' released by the JD.COM Big Data Institute, the online shopping consumer market for middle-aged and elderly people has great development potential in the context of the continuous penetration of the Internet into middle-aged and elderly people.

## 2. RESEARCH ON ONLINE SHOPPING MARKET OF MIDDLE-AGED AND ELDERLY PEOPLE

### 2.1 SAMPLING SURVEY

Since the accuracy of sampling is related to the order of sampling methods, and the more advanced the sampling method used, the greater the proportion of sampling accuracy, in order to ensure the scientific nature of the survey, reduce errors. Firstly, the stratified sampling method was selected. In order to ensure the rationality of the sampling process, a two-stage PPS sampling was used to conduct a trial survey and a formal survey.

The formal investigation mainly adopts interview, supplemented by electronic questionnaire and in-depth interview.

The survey data collection is divided into three categories: the elderly, children, children and their parents data. Data coding using both pre-coding and post-coding to achieve two-way entry of the questionnaire, each questionnaire questions in turn. The reliability and validity of the questionnaire data and the project test are carried out. All the tests of the questionnaire are passed. The questionnaire structure is reasonable, and the survey results are reliable.

### 2.2 STATUS OF MIDDLE-AGED AND ELDERLY SURVEYS

Using descriptive statistical analysis correspondence analysis, ordered multi-classification Logistic regression and correlation analysis, this paper analyzes the subjective and objective factors that affect the willingness to participate in online shopping, the satisfaction of participating in online shopping, and the subjective and objective factors that hinder the participation of middle-aged and elderly people, and longitudinally compares the development status of middle-aged and elderly markets in online shopping platforms, and makes deep processing of

data to provide data support for exploring how to further establish media platforms.

Through K-means++, the middle-aged and elderly groups are subdivided and user portraits are depicted. Through the market survey, the conclusions are as follows: (1) The development potential of middle-aged and elderly online shopping market is large, and the attention of social support is high ; (2) The participation of middle-aged and elderly people in online shopping is limited to their original living habits, and subsequent companionship or increased participation ; (3) The satisfaction of middle-aged and elderly people to online shopping needs to be improved, and the restrictive factors are mainly e-commerce platforms ; (4) Online shopping reflects the social ' digital divide ' problem, the aging society humanistic care is urgent ; (5) There are differences in online shopping preferences among middle-aged and elderly groups, but the main shopping types are daily necessities, food and clothing.

### 3.DEVELOPMENT OF AGING PLATFORM

Based on the above research results of online shopping for middle-aged and elderly people, combined with the theoretical basis of literature, this paper compares the consumption demand of middle-aged and elderly people in the social environment, the market situation of middle-aged and elderly people in the e-commerce platform, and analyzes and compares the contradiction between supply and demand of middle-aged and elderly modules in the e-commerce platform to explore the best shortcut to improve the participation of middle-aged and elderly people in online shopping. Through sorting out the ideas and conclusions of relevant literature, further processing of data, and selecting constructive suggestions for research, this paper provides theoretical support for exploring how to further establish media platforms.

Based on the characteristics of the middle-aged and elderly consumer groups, this paper discusses the feasibility of combining intelligent recommendation and speech recognition to optimize the shopping page. Combined with the existing excellent cases and strong theoretical basis, the middle-aged and elderly groups in the data are subdivided through K-means++, and the portraits of users are portrayed. The data processing and analysis are attempted, and the flow pool is set up for APP development. The connection with relevant platform channels is attempted to integrate information and promote the R & D landing.

Based on the current situation of online shopping market, through targeted research, this paper explores the specific improvement measures of e-commerce platform, and tries to take the form of lap, screen products with high cost performance from each e-commerce platform, and present them to the middle-aged and elderly according to the principle of font amplification and simple interface after extracting key information of products.

The established APP pre-model is compared with other similar APPs to establish a feedback optimization mechanism. First of all, a small range of applicable, for the promotion of the market is not suitable for non-standard settings to optimize and upgrade. Always adhere

to the principle of aging first, and absorb and adopt the relevant suggestions put forward by users. Follow the front end of the industry, grasp the general direction, and strive to promote the media to deeper areas.

### 3.1RESEARCH ON THE MARKET DEVELOPMENT OF MIDDLE-AGED AND ELDERLY ONLINE SHOPPING

Through descriptive statistical analysis of the basic information of the respondents, it is found that both urban and rural middle-aged and elderly people have a certain understanding of online shopping. Limited by economic level and social infrastructure, rural online shopping is slightly lower than urban. But for the online shopping of middle-aged and elderly people, the attention and support of the whole society show a ' one-sided ' trend, which reflects the public ' s affirmation of the online shopping trend of middle-aged and elderly people. Most middle-aged and elderly people have the intention of online shopping. Based on this phenomenon, the intervention of large factories and investment capital will increase the development potential of the online shopping market of the ' silver hair ' group. In addition, more middle-aged and elderly people buy daily necessities, clothing, food, household appliances online, pay attention to practicality. This requires the standardized operation of e-commerce platforms to ensure the commodity quality of the online shopping market for middle-aged and elderly people, so as to stimulate market vitality. Through the establishment of the Spearman model, it can be seen that among the influencing factors of online shopping satisfaction proposed by middle-aged and elderly people, they are more inclined to online shopping platform issues, such as page complexity, after-sales service attitude, and concerns about counterfeit and inferior products. E-commerce platforms and society should make relevant measures in this regard.

### 3.2FEASIBILITY STUDY ON E-COMMERCE PLATFORM AND MEDIA DEVELOPMENT FOR MIDDLE-AGED AND ELDERLY PEOPLE

Through the comprehensive analysis of the main influencing factors of online shopping for middle-aged and elderly people, it is concluded that the optimization of shopping page by combining intelligent recommendation and speech recognition will greatly improve the enthusiasm and satisfaction of middle-aged and elderly consumers in online shopping. The simple and easy-to-learn platform will become an important link for middle-aged and elderly people, a typical consumer to participate in online shopping. In short, improving the feasibility of deep development of media platform research and development will greatly improve the scale economic benefits of middle-aged and elderly markets.

With the development of aging society, the whole society has paid more attention to the life of middle-aged and elderly people, and the emergence of the epidemic has increased the market quota of online shopping. The middle-aged and elderly consumer market ushered in an important period of strategic opportunities. Therefore, the in-depth development and innovation of the online shopping middle-aged and elderly market not only solves

the existing contradiction between supply and demand, but also injects vitality into economic development. It can better adapt to the current social development trend.

### 3.3 PLATFORM TECHNICAL SCHEME

In order to make the middle-aged and elderly people better integrate into the online shopping market, the project constructs a shopping platform exclusive to the middle-aged and elderly people -- goods APP. In cooperation with Taobao, Jingdong, Pinduoduo and other platforms, the product information is focused on screening, in order to expand the font, the page is concise and clear, and the page is displayed, with voice input and voice playback functions. The search content is double audited to prevent the elderly from being cheated, and the flow pool intelligent recommendation system is designed to recommend the top-ranking items to the elderly and reduce the selection of shopping. Voice input and playback functions, dual audit and flow pool intelligent recommendation system technology have been realized in Douyin and other APPs. So the technology of this project is scientific and feasible.

Based on the above analysis, it shows that the research of this project is feasible in terms of content, method and implementation. APP part of the page is as follows. As shown in Fig. 1.

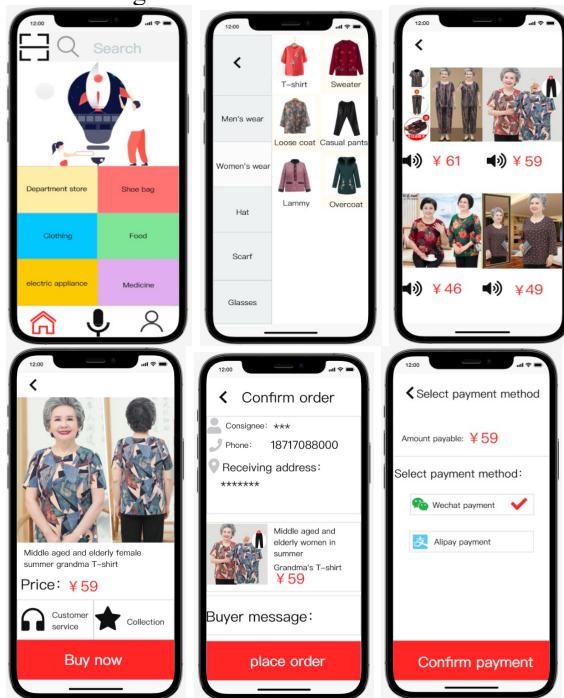


Figure 1: APP section page

### 4.PURPOSE AND SIGNIFICANCE

Under the trend of aging development, focusing on the middle-aged and elderly economy, starting from the contradiction between online shopping supply and demand of middle-aged and elderly people, this paper collects the understanding of online shopping of middle-aged and elderly people, analyzes the online shopping preferences of middle-aged and elderly people, understands the favorite online shopping products of middle-aged and elderly people of different ages, analyzes

the willingness of middle-aged and elderly people to participate in online shopping, and studies the factors affecting the participation of middle-aged and elderly people in online shopping, so as to find ways to improve the willingness of middle-aged and elderly people to participate in online shopping, and provide development direction for middle-aged and elderly e-commerce platforms, so as to create a more suitable online shopping environment for middle-aged and elderly people. From the perspective of user experience, this paper analyzes the shortcomings of online shopping, and provides some reference for the future development direction of online shopping platform, in order to promote the development of middle-aged and elderly online shopping market. Solve the contradiction between supply and demand at the same time stimulate the vitality of the elderly economy.

### REFERENCES

- [1] Han Zhenqiu. Sustainable Development Strategies for China's Aging Society [J]. *Guangxi Social Science*, 2019 (03): 86-89.
- [2] Dong Xiaoni, Zhu Wenxiu. Supply side design and optimization of the elderly e-commerce market from the perspective of long tail theory [J]. *Science and technology economic journals*, 2019, 27 (14): 23-24.
- [3] Li Zhenzhen. Research on the network consumption behavior of urban elderly groups in China [D]. Huaqiao University, 2014.
- [4] Luo Weichen. Analysis of middle-aged consumer behavior under the background of 'Internet +' [J]. *China's strategic emerging industries*, 2017 (28): 49-50.
- [5] Yang Rui, Chen Suyue, Liu Shasha, Li Fei, Chen Jiawen. Analysis of online shopping behavior of elderly groups and marketing suggestions of e-commerce platform - Taking Nanjing as an example [J]. *China market*, 2020 (12): 187-188.
- [6] Zhang Yiqi. Middle-aged and elderly people embrace online shopping [J]. *Finance: Thesis*, 2018 (4): 33-34.
- [7] Talk about Jiang. Don't let the 'Double 11' national carnival only become 'the carnival of young people' [N]. *Consumer Daily*, 2018-11-14 (a04).
- [8] Peng Renfu, Huang Chengju. Research on the construction of ecological benign development mechanism of online shopping for the elderly based on cdd concept [J]. *Modern marketing*, 2019 (12): 212-213.
- [9] Chu Junhong. Elderly consumer market services in the United States and France [J]. *Market and demographic analysis*, 1995, (04): 56-57.
- [10] Li Wei. Under the background of population aging, where does network consumption go Based on the analysis of the current situation of the online consumer market of middle-aged and elderly products [J]. *Lanzhou Academic Journal*, 2017 (03): 191-199.
- [11] Zhang He. Research on the elderly consumption market in China under the background of population aging [D]. Southwest University of Finance and Economics, 2011.
- [12] Zhang Zixuan. Get the elderly on the 'Internet+' express [J]. *Good evening*, 2020 (10): 17-18.

# Evaluation of Higher Education System based on index Dimension Reduction

Mengyuan Yang<sup>1</sup>, Zihang Chen<sup>2</sup>, Bingbing Shi<sup>3</sup>

<sup>1</sup>College of Science, North China University of Science and Technology, Tangshan 063210, China;

<sup>2</sup>School of Artificial Intelligence, North China University of Science and Technology, Tangshan 063210, China;

<sup>3</sup>School of Artificial Intelligence, North China University of Science and Technology, Tangshan 063210, China

**Abstract:** Education industry is an important part of the service industry. Due to the uniqueness of national higher education system, the research and evaluation of a country's higher education system has become a hot spot. Therefore, a model is established to solve these problems. For the model, firstly, the relevant indicators of China, the United States, Germany, Vietnam and higher education are collected, and the framework of the evaluation model is listed: Topsis model based on entropy weight method, and the improved countries are selected; Principal component dimension reduction analysis and factor dimension reduction analysis further concentrate indicators; Topsis model based on entropy weight method was used to evaluate the health degree of the system.

**Key words:** Entropy Weight Method; Principal Component Analysis; Factor Analysis; Evaluation Of Education System

## 1.PROBLEM ANALYSIS

This topic requires the development of a model that can be used to assess the health of higher education systems in any country, and explores the establishment and vision of education systems. Our work mainly includes the following aspects [1-4]:

Based on the data collected in four countries, the model of education evaluation system is established, and a national higher education system with room for improvement is selected. Based on principal component analysis, the known data indexes were combined into several unrelated indexes, and the newly established indexes were named.[5] Finally, the corresponding analysis and suggestions were put forward. Based on factor analysis, the index data of selected countries were analyzed and the above analysis was further verified. Based on the above model, the influential indicators were found through factor analysis method, and the influence of the factors on the education system was properly analyzed according to the influential factors.[6] In addition, the entropy weight method and Topsis were combined to establish a model for evaluation, including the implementation of effective measures to represent the indicators and increase investment.

## 2.MODEL

### 2.1DATA PROCESSING

First, select the four countries, namely China and Vietnam, the United States, Germany, and then select the eight indicators associated with higher education system, namely, universities, education spending as a percentage of GDP[7-8], undergraduate course graduation rate, the development of higher education levels, number of PhD

graduates, students and graduates employment, proportion of university teachers and students, QS number in the top 100 university.

### 2.2TOPSIS MODEL BASED ON ENTROPY WEIGHT METHOD

(1) Identify a range of factors related to the higher education system

There are many factors affecting the higher education system. Firstly, eight representative indicators are selected: the number of universities, the proportion of education expenditure in GDP, the undergraduate graduation rate, the number of doctoral students, the number of international students, the graduate employment rate, the proportion of teachers and students to universities, and the number of top 100 universities in QS. If these five indicators are set as evaluation factors, the evaluation factor is set as  $X_j=\{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$

(2) Assessing higher education systems in different countries

Eight indicators are used to compare and measure a country's higher education system. Therefore, the following Table 1 is obtained:

The evaluation model	Indicators of higher education systems							
	Universit y numbe r	Educatio n spendi ng as a percentag e of GD P	Undergra duate gr aduation Rate	Number of PhD graduates	Number of stude nts studi ng abroa d	Graduate employm ent rate	Percenta ge of co llege stu dents an d faculty	Number of univer sitier in the top 100 rank ed by Q S
China	2914	4.04	95	62578	140637	93.5	17.6	6
USA	3621	7.3	60	55195	1075496	82	14.41	27
Germany	426	4.9	34	29000	282000	91	12.3	3
Vietnam	235	5.7	96.33	24000	13758	38	20.28	0

(3) Standardized data processing

The original matrix of education system state can be expressed as evaluation factor subset  $X_j = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$  And higher education systems in different countries:

$$A = \begin{pmatrix} a_{11} & \cdots & a_{18} \\ \vdots & \ddots & \vdots \\ a_{81} & \cdots & a_{88} \end{pmatrix}$$

Due to the size difference, the initial values in the original matrix are difficult to compare, so these indicators need to be standardized. The processed standard matrix is:

$$B = \begin{pmatrix} b_{11} & \cdots & b_{18} \\ \vdots & \ddots & \vdots \\ b_{81} & \cdots & b_{88} \end{pmatrix}$$

(4) Calculate the information entropy of each index

According to the calculation formula of information entropy:

$$S_j = -k \sum_{j=1}^n f_{ij} \ln f_{ij}, j = 1, 2, 3, 4, 5$$

$$f_{ij} = \frac{b_{ij}}{\sum_{i=1}^m b_{ij}}, k = \frac{1}{\ln n}$$

When  $f_{ij} = 0$ ,  $f_{ij} \ln f_{ij} = 0$ , so as to obtain the information entropy vector of each evaluation index.

(5) Calculate the weight of each evaluation index

Set the weight vector of each evaluation index as  $W = (w_1, w_2, w_3, w_4, w_5)$ . According to the calculation formula of index weight:

$$w_j = \frac{1 - E_j}{m - \sum_{j=1}^m E_j}$$

### 2.3 PRINCIPAL COMPONENT ANALYSIS

Suppose there are  $n$  samples and  $P$  indicators, then the sample is  $n \times P$  matrix  $X$  is:

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix}$$

You need to find a new set of variables  $z_1, z_2, z_3, z_4, z_5$ . Then they need to meet:

$$\begin{cases} z_1 = l_{11}x_1 + l_{12}x_2 + \cdots + l_{1p}x_p \\ z_2 = l_{21}x_1 + l_{22}x_2 + \cdots + l_{2p}x_p \\ z_3 = l_{31}x_1 + l_{32}x_2 + \cdots + l_{3p}x_p \end{cases}$$

Analysis steps:

(1) Standardized sample matrix. (2) Calculate the covariance matrix of standardized samples; (3) Calculate the eigenvalue and eigenvector of R. (4) Calculate the principal component contribution rate and cumulative contribution rate. (5) Write down the principal components and analyze their meaning.

### 3. SOLUTION OF MODEL

#### 3.1 SELECTION OF COUNTRIES TO BE IMPROVED

Through the analysis of the evaluation model, Vietnam is finally selected as the country to be improved.

#### 3.2 REANALYSIS OF IMPROVED COUNTRIES

Using the indicators after principal component analysis, the evaluation model is used again to make a reasonable evaluation of Vietnam, and the factor analysis method is used to select the impact indicators that have a great impact on Vietnam, and put forward relevant suggestions to improve the education level of Vietnam.

### 4. MODEL TEST

#### 4.1 FACTOR ANALYSIS TEST

According to kmo and Bartlett test of SPSS:

Tab 2: KMO and Bartlett tests		
Kmo sampling suitability quantity	0.912	
Bartlett sphericity test	Approximate chi square	709.002
	Degrees of freedom	26
	Significance	0.000

(1) As shown in Tab 2: Kmo test:  $kmo > 0.9$  indicates that it is very suitable for factor analysis.

(2) As shown in Tab 2: Bartlett spherical test: where the p value corresponding to the statistical data is less than 0.05 (95% confidence level), reject the null hypothesis, and consider that the correlation coefficient cannot become the identity matrix, that is, there is correlation between the original variables, which is suitable for factor analysis.

### REFERENCES

- [1] Yang Hongkun. Exploration and Research on course certificate integration in Higher Vocational Education [J]. Automobile maintenance technician, 2021 (10): 126-127.
- [2] Zhang Yong. Progress, challenge and path choice of education evaluation system construction in China [J]. Journal of Hebei Normal University (EDUCATIONAL SCIENCE EDITION), 2021, 23 (05): 17-23.
- [3] Liu Fei, Yin Haiqiang. Research on the construction of formative evaluation system under the background of higher education quality revolution [J]. Science and education guide, 2021 (25): 4-6.
- [4] Niu Lu Shuai, Peng Fu. Research on real-time movie recommendation algorithm under big data platform [J]. Software engineering, 2021, 24 (09): 13-16.
- [5] Kang Zhuang, Yang Hao. Application of big data platform based on Hadoop architecture in power generation enterprises [J]. China Science and technology information, 2021 (17): 43-44.
- [6] Zhang Haifeng, Wei Kexin. A query method based on spark big data processing platform [J/OL]. Journal of Nanjing University of Posts and Telecommunications (NATURAL SCIENCE EDITION), 2021 (04): 82-90 [2021-10-18].
- [7] Yang Qihui. Thoughts on the development of China's Higher Education under the new situation of globalization [J]. Educational theory and practice, 2021, 41 (24): 8-12.
- [8] Zhou shaoke, Wang Lei, Cui Lin, Guo Xuan, Wan Xiaozhou. Fully distributed cluster deployment of big data Hadoop technology [J]. Industrial control computer, 2021, 34 (08): 101-103.

# Research on Value Characteristics of Consumer Behavior Based on Factor Analysis

Ziyang Yu<sup>1\*</sup>, Xiaodie Zhuang<sup>2</sup>, Jing Liu<sup>3</sup>

<sup>1</sup>Mathematical Modeling Association, North China University of Science and Technology, Tangshan, 063210, China;

<sup>2</sup>College of science, North China University of Science and Technology, Tangshan, 063210, China;

<sup>3</sup>Institute of Mining Engineering, North China University of Science and Technology, Tangshan, 063210, China;

\*Corresponding Author.

**Abstract:** With the continuous development of the Internet, companies in various fields pay more and more attention to expanding the channels of Internet customers. A fast and effective model is needed to introduce fresh and active users, improve users' desire to buy products and enhance the company's. The purpose of brand influence. However, how to identify high-quality users and channels and optimize marketing costs has always been a pain point for companies. Based on this, this paper deals with user information through data preprocessing, data visualization, factor analysis and multiple linear regression. Method, such as analysis of user behavior data. Then judge the value of users, and make special marketing strategies for users. Now, we are promoting at a low cost to achieve the goal of improving the conversion rate of users.

**Keywords:** Multivariate linear regression; TOPSIS scoring factor analysis of consumer behavior.

## 1. INTRODUCTION

In recent years, the popularity of the Internet has brought earth-shaking changes to the traditional marketing concept in the market. Their strong competitiveness is actively expanding the channels of obtaining customers from the Internet<sup>[1-2]</sup>, thus introducing fresh and active uses for the company's products from various aspects. Users, improve the user's desire to buy, enhance the company's brand influence. Enterprises constantly adapt to the development of the times, the traditional camp. Marketing ideas and strategies to improve, to further tap the useful information for themselves, and keep continuous learning and diligence. Ability, looking for the best customer channels. However, it has been a long time for households to identify high-quality users and channels and optimize marketing costs. Can't be solved effectively. In this paper, we need to judge the value of users by analyzing the given user behavior data. Users are divided into important users, ordinary users and development customers according to certain criteria. And then use for different categories. Make special marketing strategy, realize low-cost promotion, and reach the standard of improving user conversion rate.

## 2. MODEL BUILDING

First of all, in order to avoid the influence of index dimension, we get three grades by combining the scoring criteria. Level of index set, standardize the original data to the same order of magnitude, and then comprehensive comparative analysis. In addition, after we do the positive

treatment on the factors, we standardize the three factors, so as to get more scientific comprehensive score data.

### 2.1 DATA PREPROCESSING

Data preprocessing includes: removing unique attributes, data standardization, normalization, feature selection (dimension reduction).

(1) Remove unique attributes: Unique attributes are usually ID attributes, which can't describe the distribution law of the sample itself. Here, the ID attribute represents the basic attribute characteristics of users. We need to determine the categories of users with different IDs, so the ID cannot be removed, but we can remove the two unique attributes such as equipment and mobile phone model. At the same time, because the login days cannot be negative, we deleted the data of -1 login days.

(2) Standardized data processing: First, in order to avoid the influence of index dimensions, we get three equal-grade index sets by combining with score evaluation criteria, standardize the original data to the same order of magnitude, and then comprehensively compare and analyze. In addition, after we have done the positive processing on the factors, we have standardized the three factors to get more scientific and reasonable comprehensive score data.

(3) Feature selection (dimension reduction): We do factor analysis on the data in login\_day except user id, attention to WeChat official account 1, attention to WeChat official account 2, adding sales friends and joining the group, and finally get three factors, in order to get the comprehensive score of each user, and classify users according to certain criteria.

(4) Positive data processing: As a result, three factors are finally obtained, the first factor is called the number of study courses, the second factor is called purchase desire, and the third factor is called login interval. Since the login interval is a very small index, it needs to be positively processed and converted into a very large index. The formula for converting a very small index into a very large index is as follows:  $\tilde{x}_i = \max - x$

### 2.2 KMO\BARTLETT'S TEST

In this paper, KMO and Bartlett were used to test the raw data to verify whether the raw data could be applied to the factor analysis model. The test results are shown in Table<sup>[3-4]</sup>.

Table 1 KMO and Bartlett's test

Sampling suitability quantity of KMO.	0.716
	Approximate chi-square 398871.722
Bartlett sphericity test	Freedom 55
	Significance 0.000

The test results showed that factor analysis could be performed using the original data at 95% confidence intervals.

### 2.3R-TYPE FACTOIE MODELING

Constructing the relationship between the 11 indicators and the common factor yields the following determinant.

$$\left\{ \begin{array}{l} x_1 = u_1 + a_{11}f_1 + a_{12}f_2 + \dots + a_{1m}f_m + \varepsilon_1 \\ x_2 = u_2 + a_{21}f_1 + a_{22}f_2 + \dots + a_{2m}f_m + \varepsilon_2 \\ \dots \\ x_{11} = u_{11} + a_{111}f_1 + a_{112}f_2 + \dots + a_{11m}f_m + \varepsilon_{11} \end{array} \right.$$

$a_{ij}$  denotes the covariance between the origin alvariable  $x_i$  and the common factor  $f_j$ ,  $a_{ij} = \text{cov}_{ij}(x_i, f_j)$ . Extracting  $a_{ij}$  from the determinant constructs the A matrix. The whole determinant can be abbreviated as:  $x = \mu + Af + \varepsilon$

where  $f = (f_1, f_2, \dots, f_m)$  ( $m \leq 11$ ) is the common factor vector,  $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m)$  is the special factor vector,  $A_{s \times m} = (a_{ij})$  is called the factor loading matrix, and the rank of the A matrix is assumed to be m.

The dependence of the original variable  $x_i$  on the common factor and the contribution of the common factor  $f_j$  to the overall index data are solved by calculating the sum of squares of row elements  $h_i^2$  and the sum of squares of column elements  $g_i^2$  of the A matrix, respectively, and then the most dominant composite factor is selected.

where the sums of squares of the row and column elements are, respectively:  $h_i^2 = \sum_{j=1}^m a_{ij}^2 g_i^2 = \sum_{i=1}^p a_{ij}^2$

### 2.4 DETERMINE THE NUMBER OF FACTOES

Three public factors were extracted by the cumulative contribution of the variance interpretation table and the turning point of the gravel plot, and the variance contribution of the three extracted public factors were 30.839%, 29.029%, and 11.458%, respectively, and the cumulative variance contribution of the first three public factors had reached 71.326%, that is the first three public factors could contain 71.326% of the information of the original index, so the first three public factors were chosen. The first three common factors were chosen to reflect the data of the overall index.

### 2.5 FACTOR SCORE

Factor analysis is to represent the variables as linear combinations of public and special factors, and in this paper, we can reverse the public factors as linear combinations of the original variables to obtain the factor scores, and the generated component score coefficient matrix is shown in Table 2.

**Table 2 Component score coefficient matrix**

	ingredient	1	2	3
login day	0.241	0.079	0.261	
login diff time	0.007	0.229	0.482	
distance day	0.031	0.027	0.590	
course order num	0.076	0.445	-0.311	
finish num	0.279	-0.188	0.001	
coupon	0.076	0.464	-0.249	
study num	0.185	0.025	-0.069	
login time	0.153	-0.123	-0.209	
learn num	0.288	-0.190	-0.002	
camp num	0.107	-0.185	-0.193	
launch time	0.157	0.322	0.147	

From the above table, the expressions of the three principal components can be obtained.

$$f_1 = 0.241x_1 + 0.007x_2 + 0.031x_3 + \dots + 0.157x_{11}$$

$$f_2 = 0.079x_1 + 0.229x_2 + 0.027x_3 + \dots + 0.322x_{11}$$

$$f_3 = 0.261x_1 + 0.482x_2 + 0.590x_3 + \dots + 0.147x_{11}$$

The first principal component  $f_1$  has a larger coefficient of positive loadings for the number of study sessions and the number of completed sessions, so the first principal component is said to be the number of studies; the second principal component  $f_2$  has a larger coefficient of positive loadings for the number of coupons received and the number of unfinished orders with annual classes, so the second principal component is said to be the desire to purchase; the third principal component  $f_3$  has the largest coefficient of positive loadings for the logging interval and the number of days between the last logging and the end of the period, so the third principal component is said to be the study interval. The three principal components are more intuitive to determine whether users will eventually place orders from three perspectives: learning quantity, purchase desire, and learning interval.

### 3.CALCULATION OF TOPSIS SCORE

Through the analysis of the three principal component indicators, it can be seen that the learning quantity and purchase desire are the maximum indicators, while the learning interval is the minimum indicator. Set up the scoring model based on TOPSIS.

#### 3.1 FORWARD TREATMENT

As the learning interval is a minimal index, we need to forward the minimal index and transform it into a maximal index<sup>[5-6]</sup>. The formula for transforming the minimal index into a maximal index is as follows:  $\tilde{x}_i = \max - x$

Among them,  $\tilde{x}_i$  is the index data after the very small index is turned forward,  $x$  is the original data, and  $\max$  is the largest value among similar indexes.

#### 3.CALCULATE THE SCORE AND NORMALIZE IT

Define the distance between the i-th ( $i = 1, 2, \dots, n$ ) evaluation object and the maximum value and the minimum value as follows:

$$D_i^+ = \sqrt{\sum_{j=1}^n w_j (z_j^+ - z_{ij})^2}$$

$$D_i^- = \sqrt{\sum_{j=1}^n w_j (z_j^- - z_{ij})^2}$$

The TOPSIS score corresponding to whether the user places an order is:  $S_i = \frac{D_i^-}{D_i^+ + D_i^-}$

### 4.EVALUATION MODEL OF MULTIPLE LINEAR REGRESSION

#### 4.1 TEST OF SIGNIFICANCE

To test whether there is a real linear relationship between the dependent variable and the independent variable<sup>[7-8]</sup>, the significance test is carried out first:

Test of significance	model	Model fitting condition	Likelihood ratio test
----------------------	-------	-------------------------	-----------------------

	-2 log likelihood	chi-square	freedom	significance
finally	0.000	99.779	99	0.0015

P=0.0015<0.05, that is, within 95% confidence interval, the regression equation is considered meaningful, and the model is effective.

#### 4.2 USING MULTIPLE LINEAR REGRESSION METHOD

The following multiple linear regression is used to predict each index coefficient. Let three principal components be independent variables and  $y$  be dependent variable, and satisfy the linear relationship as follows:

$$y_i = \beta_0 + \sum_j B_j f_{ij} + \mu_i,$$

$$i = 1, 2, \dots, n \quad j = 1, 2, \dots, p$$

Where  $\beta_0, \beta_1, \dots, \beta_p$  is the regression coefficient,  $\mu_i$  is an

model	Unstandardized coefficient		Normalization coefficient	t	significance
	B	Standard error	Beta		
1	(constant)	0.140	.000	93085513	0.000
	x	0.287	.000	474520772	0.000
	y	0.287	.000	344314503	0.000
	z	0.287	.000	187352929	0.000

After solving the index coefficient, the regression equation is as follows:

$$y = 1.945f_1 + 1.664f_2 + 0.761f_3$$

From the above-mentioned factor model, multiple linear regression evaluation model and TOPSIS score, we can well understand that our model has a good fitting effect. Based on this, we tested the sensitivity of the model, and further fitted the result of whether the user would place an order with the given data, and judged that the fitting effect of the model reached 95%, indicating that the fitting effect of the model was good.

#### 5. EVALUATION OF MODEL

- (1) In order to evaluate whether the user finally places an order or not, we divide the influencing factors into three categories through factor analysis and dimension reduction, which makes the model simpler.
- (2) The use of multiple linear regression combined with TOPSIS comprehensive evaluation when weighting each index avoids the influence of individual subjective factors.
- (3) In this paper, we consider the influence of many factors when solving whether the user finally places an order, which makes the established model more reliable and applicable.

#### 6. GENERALIZATION OF THE MODEL

This model is applicable to a wide range of fields, realizes the overall analysis of whether users ultimately buy under the influence of multiple factors, can be applied to a variety of sales problems, and plays an extremely important role in improving users' desire to buy, optimizing marketing costs and distinguishing high-quality users.

#### 7. CONCLUSIONS AND SUGGESTIONS

According to the research of users' consumption behavior, it is found that whether users place orders or not is related to multiple factors. After the model is established, it is finally concluded that whether users place orders or not is closely related to such factors as watching time, active time, login interval, age, etc. Now, some suggestions are put forward for these factors.

unobservable disturbance term that satisfies certain conditions. Make the predicted value.

$$y_i = \beta_0 + \sum_j B_j f_{ij} + \mu_i,$$

$$i = 1, 2, \dots, n \quad j = 1, 2, \dots, p$$

Where  $\mu_i$  is residual, which represents the numerical difference between the actual value and the predicted value. Based on the above analysis, we define  $f_1$  as learning quantity,  $f_2$  as purchasing desire and  $f_3$  as learning interval. The multiple linear regression evaluation model of whether users place orders is constructed as follows:

$$y = \beta_0 + \beta_1 f_1 + \beta_2 f_2 + \beta_3 f_3$$

#### 4.3 CALCULATE REGRESSION COEFFICIENT

#### 7.1 CLASSIFY USERS AND GIVE DIFFERENT PREFERENTIAL TREATMENT

Login duration can be used as one of the indicators to distinguish the importance of users, so we can divide users into three categories according to the division of login duration: membership, silver and bronze. Different incentives are given to these three types of users respectively: member users enjoy full discount of products and send small gifts; Silver can enjoy the relief of large coupons; However, bronze users can only enjoy occasional coupon grabbing activities, so as to stimulate users' promotion psychology and achieve the purpose of increasing users' consumption.

#### 7.2 "MEMORY" PUSH

Most users will log in again within 0.8-1 days after logging in. The company monitors the online time of users this time, and sets the time to push the products browsed by users this time, and push similar products and related products again.

#### 7.3 GRADING, TAKING DIFFERENT MEASURES

According to the number of users, the region is divided into three levels: sales Golden Zone, sales retention area and sales barren area. Sales Golden Zone has a larger user flow. In these areas, you can put the products that users are interested in in your own golden position, put the content with the best transformation potential there, and don't waste every exposure opportunity, so as to ensure the click volume and order rate of users; As for the sales retention areas, which belong to areas with high sales flow, users in these areas are also the main objects of sales, so we should strengthen publicity and promotion, and add novel elements to give users a good visual experience and consumption experience; For poor sales areas, which belong to areas with low sales volume, this part of users has poor stickiness and low consumption contribution. Therefore, the company should reduce the product launch, design eye-catching product appearance, and initially improve users' attention.

#### 7.4 PUSH PRODUCT TYPES BASED ON AGE

The age of consumers is concentrated in the middle-aged and old age group (46-99). The company should conduct market research, collect the products with the highest sales volume in this age group, and increase the efforts of putting in the products needed by this group of people. For example, middle-aged users are busy and need to push products or methods to relieve fatigue and promote sleep; The elderly users are in poor health, so they need to push health care, health care products and some courses to strengthen their health. According to the different needs of different age groups, different types are provided.

#### 7.5 INCREASE THE PREFERENTIAL EFFORTS FOR USERS TO PURCHASE THE GOLDEN PERIOD.

At the time of gold purchase, the company can enhance the popularity of the store by means of disguised discounts, sample promotion, red envelope promotion, etc., and promote customers to place orders again in a short time, thus effectively forming customer loyalty.

Discount in disguise: postage-free. The price of a single product is slightly lower than that required for postage-free. Buy two or more products together to meet the requirement for postage-free. Combine the products for a discount.

#### REFERENCES

- [1] Zhai Jinzhi. Analysis of Internet users' consumption behavior based on big data [J]. *BusinessEconomics Research*, 2020(24):46-49.
- [2] Zheng Xin. Analysis of the influence of consumer behavior on consumer consumptionbased on Gstore data [J]. *Journal of Anyang Institute of Technology*, 2020, 19(06):46-50+109.
- [3] Yao Xin, Bian Hualin, Li Youzhi, Zhou Qiaoqiao. Analysis of carbon density and its influencing factors in the south of Dongting Lake Wetland [J/OL]. *Journal of Hunan Agricultural University (Natural Science Edition)*.
- [4] Zhao Baojun, Liu Feng, Zhang Junpei, Li Jing, Ma Qingguo, Li Dongsheng, Wang Kehan, Ma Caixia, Gong Yonghong. Factor Analysis and Comprehensive Evaluation of Yield Traits of Walnut Hybrids [J/OL]. *Molecular Plant Breeding*: 1-12[2021-10-20].
- [5] Qi HUANG. Realization Path for Inclusive Finance to Support Rural Revitalization in Poverty-stricken Areas[J]. *Asian Agricultural Research*, 2021, 13(06):1-5.
- [6] Ma shan-shui, Lin Shi-ping, Yang Chao. quantitative study on visual appreciation of tufted plants based on entropy TOPSIS method[J]. *Guangdong landscape architecture*, 2021, 43(03):55-60.
- [7] Dong xiaomeng. analysis of influencing factors of statistical professional satisfaction based onmultiple linear regression [J]. *value engineering*, 2019, 38(36):276-277.
- [8] Zheng Haipeng. Analysis of employment salary expectation of college students based on SPSS multiple linear regression [J]. *Journalof Qiqihar University (Natural Science Edition)*, 2019, 35(06):77-81.

# Photoelectric Intelligent Garbage Sorting Vehicle

Tengda Zhang<sup>1, 2</sup>, Jing Liang<sup>1, 2</sup>, Xiangliang Li<sup>1, 2</sup>

<sup>1</sup>Mathematical Modeling Innovation Lab, North China University of Science and Technology, Tangshan 063210, China;

<sup>2</sup>School of Yisheng Innovation Education North China University of Science and Technology, Tangshan 063210, China

**Abstract:** With the development of The Times and the progress of society, more and more garbage is produced. At present, most small and medium-sized garbage plants still need a large number of porters when classifying and transporting garbage. The work repeats machinery and has high labor intensity, which is suitable for mechanization improvement and reducing labor costs. The intelligent car designed in this paper has fast speed, stable driving process, and has camera recognition and intelligent grasping functions. It also uses ant colony system to realize intelligent trajectory planning of the car. It is an integrated system, which is simple, efficient and low cost, and can effectively improve work efficiency and reduce labor costs.

**Key words:** Garbage classification; Ant colony system; Intelligent vehicle

## INTRODUCTION

With the continuous expansion of city size and population, the annual growth rate of garbage output in China is more than 10 percent, and nearly 150 million tons of municipal garbage is generated every year. Although garbage classification has been promoted in China, people in many places lack the awareness of garbage classification, which also leads to more tasks in the garbage sorting link. [1] Garbage encircles one-third of China's cities, occupying 750, 000 mu of land, according to a survey released recently[2]. Garbage siege is becoming an increasingly difficult problem for Chinese cities. There is an urgent need for more efficient treatment of urban garbage in China.

Although garbage classification has been achieved in some areas in recent years and some achievements have been made, the garbage classification mode at the present stage is not precise enough. For example, kitchen garbage is still mixed with a large amount of plastic, glass, paper and other garbage, which cannot be directly processed by kitchen garbage processing equipment. In order to meet the requirements of on-site processing, These areas have to carry out artificial light secondary classification of household garbage after the first classification, in order to meet the standards of improving the purity of food waste disposal[3]. However, it is reported that this work has already been carried out in some areas, with tons of kitchen waste provided by waste recycling plants. However, due to the workload of sanitation workers, labor cost problems and declared failure. It can be seen that manual garbage classification is not feasible.

In this paper, on the spot, instead of the cumbersome system operation, convenient to safety and highly

intelligent, designed a model suitable for multiple application scenario of garbage sorting. It greatly simplifies the input of manpower and has many functions such as garbage identification, accurate grasping and intelligent route planning. Besides, it is easy to operate and high efficiency, which is of great significance to reduce the work burden of sorting workers.

## DEVICE DESCRIPTION

The car realizes functions including fast driving and turning, camera recognition, intelligent garbage capture(The schematic diagram of the manipulator is shown in Figure 1), intelligent trajectory planning and so on. So as to realize the search, identification, classification and emergency transport capacity of the car.

In order to improve the transport capacity of the car, fast action and turning function are necessary features, and also need to have accurate identification ability, grasp ability and excellent route planning, so that more garbage can be captured.

There is also a camera recognition feature that can identify the color of the car under the car, the color of the garbage in front and the color of the pile. For example, when the yellow accumulation point is set artificially, avoid being near the black accumulation point[4]. Similarly, correctly identifying the type of garbage ahead will be the basis of successful sorting; After capturing garbage, it is necessary to correctly put it into the corresponding color area to identify the color of the area in front.

Intelligent grasping of garbage requires accurate grasping of garbage and avoiding dropping and destruction of garbage. Accurate control of the force is very important. Here we use a manipulator with a bottom, so that the garbage is not easy to fall because of mistakes[5]. In addition, we set up several common garbage grasp strength, manipulator can according to the camera to identify the object after the signal through the MCU control grasp strength, to ensure that the garbage is not destroyed.

Intelligent trajectory planning enables the car to transport the garbage to the corresponding location.through a shorter distance.



Figure 1. Mechanical arm

The working process is as follows:

First, the car is sorting through own photoelectric sensors for autonomous navigation and search for, find to specify garbage gathering place, and then use manipulator grasping, a take a certain quantity of garbage, through the sensor to judge whether overload, reached after bearing limit stop scraping, then intelligent planning based on ant colony algorithm is the most reasonable route, get somewhere after they sort the garbage, Use the camera recognition function to judge the classification mark below, and then rotate the camera to identify the garbage and determine whether the garbage is consistent with the classification point. If so, the garbage will be placed to the classification point; if not, the next garbage category will be judged[6]. Then drive to the next classification point according to the planned route, until the garbage in the carrying device is put into place, return to the garbage collection point for the next garbage capture, repeat this process until the garbage classification point is completely sorted.

**TECHNICAL ROUTE AND FEASIBILITY ANALYSIS**  
First quick action ability need high performance drive, only a strong ability to drive can ensure maximum motor, besides must have the corresponding tire strengthen resistance to guarantee the tyres grip, not body occurred in the process of high-speed rollover and the phenomenon of slippage, car body set to streamline reduce air resistance to accelerate the car driving, make the tuyere is blunt motor, To achieve the purpose of cooling the motor.

The recognition function of the camera can adopt the computer imaging function, and the OpenMv camera and STM32 chip can be used to solve the problem of identifying the garbage and the color of the garbage accumulation place[7]. The color sensor module can also be installed at the bottom of the car to identify whether it is the pile place of the team. OpenMv module is an open source, low-cost, powerful and advanced machine vision module. With STM32H7 as the core and OV7725 camera chip integrated, the core machine vision algorithm is efficiently implemented in C language and Python programming interface is provided on the small hardware module. Machine vision algorithms on OpenMV include color block search, face detection, eye tracking, edge detection, marker tracking, etc. It can be used to detect illegal intrusion, screen defective products, track fixed markers and so on. Users can easily complete a variety of machine vision-related tasks by writing simple Python code.

Garbage grabbing device is in the front of the car, when it is determined that there is garbage in front, you can use the two small claws with a thin bottom to grab garbage, which can well avoid garbage damage and falling.

The path planning of garbage can also be designed through OpenMv camera and STM32 chip. Excellent track planning can make the car get the garbage first, which can be used to make the car quickly find the garbage and put it to the designated place in the actual production and application of the project. Here, the project plans to use ant colony algorithm to optimize the driving track of the car.

Ant colony is a very common insect colony, and the birth ACADEMIC PUBLISHING HOUSE

of ant colony algorithm is inspired by the behavior of ants foraging. In the process of crawling, ants will release a pheromone substance and form a pheromone track[8]. Ants can determine the next direction of foraging through the pheromone on the ground. When ants find a direction with a high concentration of pheromone during foraging, they will consciously forage for food in the direction with a high concentration of pheromone.

The proposed ant colony algorithm and related research process are as follows: first observe the behavior of real ants in the foraging process in nature, then build the behavior model of real ants in the foraging process, and finally use these behavior models to design the relevant algorithm. These algorithms can solve swarm and optimization problems. Different ant colony algorithms can be derived from different ant behaviors. The relation between ant behavior and ant algorithm is shown in Figure 2:

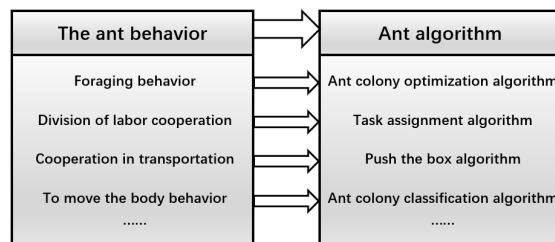


Figure 2. Diagram of ant behavior and ant algorithm  
THE KEY PROBLEM

Garbage trajectory planning is the most critical step in this project, which needs to be simulated and realized by ant colony algorithm.

Here are the basic contents of ant colony algorithm:  
Define parameters

Set the ant number  $K$ , expected heuristic factor  $\beta$ , pheromone heuristic factor  $\alpha$ , pheromone volatilization coefficient, the maximum number of iterations is set as  $N_{\max}$ , starting point  $S$ , target point  $E$ , starting taboo table  $tabuk$  as empty set, and all ants start from the starting point.

#### Probability selection

The concentration of pheromone determines the direction of the ant's next movement in the process of searching the optimal path. Suppose that at time  $t$ , the probability of ant from position point  $i$  to point  $j$  is  $p_{ij}^k(t)$ :

$$p_{ij}^k(t) = \begin{cases} \frac{[\tau_{ij}(t)]^\alpha \cdot [\eta_{ij}(t)]^\beta}{\sum [\tau_{ij}(t)]^\alpha \cdot [\eta_{ij}(t)]^\beta} & \text{if } j \in J_k(i) \\ 0 & \text{else} \end{cases} \quad (1)$$

In Formula (1),  $J_k(i)$  is the set that the ant at position point  $i$  can select for all the following paths;  $\tau_{ij}(t)$  is the pheromone concentration from point  $i$  to point  $j$  at time  $t$ ;  $\eta$  is the heuristic factor, and its size is the reciprocal

$$\eta_{ij}(t) = \frac{1}{d_{ij}} \text{ of the distance between point } i \text{ and point } j$$

( $d_{ij}$  is the distance length between point  $i$  and point  $j$ );  $\alpha$  and  $\beta$  are heuristic factors of pheromone and expectation respectively[9].

#### Path selection and saving

When each ant transfers the next node  $j$  from node I, node  $j$  needs to be added to tabu table  $tabuk$ . Then step (1) of Formula (1) is repeated. When all ants reach the end point, the cycle is ended and the path length of each ant is stored.

#### Pheromone update

After all ants have completed a path search, the individual ant paths are then updated with pheromones. The process of pheromone volatilization is set as numerical variation of volatilization coefficient  $\rho$  ( $0 < \rho < 1$ ). Pheromone update equation is as follows:

$$\tau_{ij}(t+1) = (1 - \rho)\tau_{ij}(t) + \Delta\tau_{ij}(t) \quad (2)$$

$$\Delta\tau_{ij}(t) = \sum_{k=1}^m \Delta\tau_{ij}^k(t) \quad (3)$$

$$\Delta\tau_{ij}^k(t) = \begin{cases} Q/L^k & \text{if } k \text{ chooses } i \text{ to } j \\ 0 & \text{else} \end{cases} \quad (4)$$

In Formula (2), the pheromone increment of the distance from node  $i$  to node  $j$ ; In Formula (3),  $\Delta\tau_{ij}^k(t)$  represents the pheromone left by the KTH ant in the distance from node  $i$  to node  $j$ . In Formula (4),  $Q$  is a constant, referring to the total amount of pheromones left by ants after the whole path search.  $L^k$  Represents the path length of the KTH ant in this circular search[10].

#### (5) Iteration cycle

After updating pheromones in each cycle, all ants were relocated to the starting position, and the next round of path planning began. After  $N$  iterations, an optimal path was selected from all paths.

(6) The flow of traditional ant colony algorithm is shown in the figure 3:

So that's the basic ant colony algorithm.

In this case, ant colony algorithm can be used to carry out intelligent planning for the trajectory of the car, so as to achieve intelligent planning for the trajectory of the car[11].

#### Features and innovations

#### Technical advantages

The smart car only needs a very small space to sort and treat garbage, and can work smoothly in various garbage plants to improve efficiency[12]. The intelligent car has fast speed, stable driving process, also has the camera recognition, intelligent grab function, but also through the use of ant colony algorithm to optimize the travel path, is an integrated system, the system is simple, efficient, low cost, can effectively improve work efficiency, save manpower.

#### Price advantage

In the study of the trolley in this question, in order to improve the application value of the trolley, and in consideration of some costs in actual industrial operation, the objects we use have higher cost performance[13]. Each module and quantity of garbage sorting vehicle is shown in table 1.

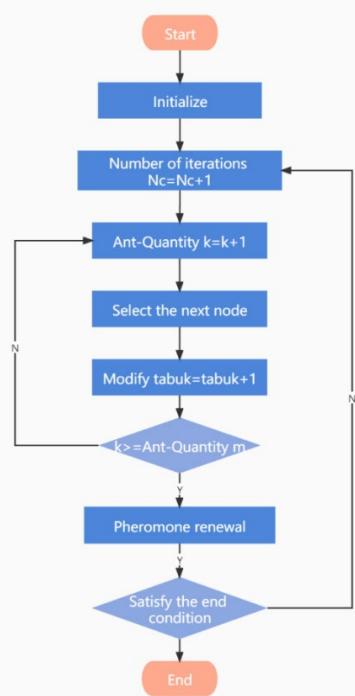


Figure 3 Flow chart of ant colony algorithm

Table 1. Each module and quantity of garbage sorting vehicle

module	quantity
OpenMV4H7Cam	1
stm32f407zgt6	1
car model	1
S3010 steering engine	1
AS1015	2
LM2940	4
XL6009	1
IR7843	8
IR2104	2
74HC02	2
MP1584	2
AMS1117	1
HCPL2630	2

In addition, ant colony algorithm has high efficiency in the application of moving cars. It belongs to bionics and is a search algorithm for finding the shortest and effective path in a specified area according to the law of ant foraging. In this application scenario, it has great advantages and conforms to the optimal trajectory of the car in the actual working environment, and can effectively reduce the economic cost caused by the potential time consumption[14].

It can be seen that whether hardware cost or running cost, the garbage sorting car designed in this paper has a very bright market prospect.

#### Algorithm advantage

Ant colony algorithm is an intelligent optimization algorithm inspired by the behavior of ants searching for

food in nature. It is based on the study of the collective foraging behavior of real ant colonies in nature, and simulates the real cooperative process of ant colonies[15]. The solution path is constructed by several ants, and the quality of the solution is improved by exchanging pheromones on the solution path.

The ant colony algorithm adopted has the following characteristics:

(1) The positive feedback mechanism is adopted to make the search process converge continuously and finally approach the optimal solution.

(2) Each individual can change the surrounding environment by releasing pheromones, and each individual can perceive the real-time changes of the surrounding environment, so individuals can communicate indirectly through the environment.

(3) Distributed computing is adopted in the search process, with multiple individuals performing parallel computing at the same time, which greatly improves the computing capacity and operation efficiency of the algorithm[16].

(4) The heuristic probabilistic search method is not easy to fall into the local optimal, easy to find the global optimal solution.

Ant colony algorithm is an excellent optimization algorithm.

## RESULTS AND DISCUSSION

Garbage is a misplaced resource. Turning waste into treasure through garbage classification can effectively relieve the situation of resource shortage[17]. However, due to various reasons in modern society, such as people's weak awareness of environmental protection, local development conditions are not allowed and various types of garbage, China's current garbage classification and treatment system is not perfect, the classification effect is not ideal, and the coverage of the area is not wide enough. For many years, most of the domestic cities are simply landfill waste treatment[18]. With the rapid improvement of social civilization, people pay more and more attention to sustainable development, while the economic growth and the improvement of people's living standards inevitably determine that human beings need to develop more resources, which will undoubtedly lead to the depletion of natural resources and the deterioration of the natural environment.

The intelligent car designed in this paper has a simple structure and perfect functions. OpenMv camera is used to accurately identify the category of garbage[19]. The manipulator with the bottom ensures the stability of grasping garbage, which can prevent the subsequent process from falling due to road inequality; In addition, we are inspired by the behavior of ants into the ant colony algorithm, which can greatly improve the driving efficiency of the garbage sorting intelligent car, greatly save the potential time cost, convenient and efficient[20]. On the one hand, garbage sorting can save land resources, and the waste of the other party is classified and sent to recycling, which can greatly reduce the occupied land. Secondly, timely disposal of garbage can effectively avoid the spread of disease and environmental pollution. Our garbage sorting vehicle realizes efficient sorting and

recycling through intelligent planning trajectory. On the other hand, garbage classification also improves the overall civilization quality of the society, helps to provide a more beautiful living environment for future generations, and helps to build ecological civilization.

## REFERENCES

- [1] De Araujo Paulo Ricardo Marques, Lins Romulo Gonçalves. Cloud-based approach for automatic CNC workpiece origin localization based on image analysis[J]. *Robotics and Computer-Integrated Manufacturing*, 2021, 68(68).
- [2] Fernandes Duarte, Silva António, Névoa Rafael, Simões Cláudia, Gonzalez Dibet, Guevara Miguel, Novais Paulo, Monteiro João, Melo-Pinto Pedro. Point-cloud based 3D object detection and classification methods for self-driving applications: A survey and taxonomy[J]. *Information Fusion*, 2021, 68(68).
- [3] Schwenzow Jasper, Hartmann Jochen, Schikowsky Amos, Heitmann Mark. Understanding videos at scale: How to extract insights for business research[J]. *Journal of Business Research*, 2021, 123(123):.
- [4] Garnett Richard L., Hanu Andrei R., Byun Soo Hyun, Hunter Stanley D.. Event selection and background rejection in time projection chambers using convolutional neural networks and a specific application to the AdEPT gamma-ray polarimeter mission[J]. *Nuclear Inst. and Methods in Physics Research, A*, 2021, 987(987).
- [5] Díaz Gastón Mauro, Negri Pablo Augusto, Lencinas José Daniel. Toward making canopy hemispherical photography independent of illumination conditions: A deep-learning-based approach[J]. *Agricultural and Forest Meteorology*, 2021, 296(296).
- [6] Wen-jing Wang, Xue-yi You. Benefits analysis of classification of municipal solid waste based on system dynamics[J]. *Journal of Cleaner Production*, 2021, 279(279).
- [7] Phuong T. Ly, Alexandra Lucas, Sio Hang Pun, Anna Dondzillo, Chao Liu, Achim Klug, Tim C. Lei. Robotic stereotaxic system based on 3D skull reconstruction to improve surgical accuracy and speed[J]. *Journal of Neuroscience Methods*, 2021, 347(prepubish):
- [8] Roberto Moscetti, Daniel Hagos Berhe, Mariagrazia Agrimi, Ron P. Haff, Peishih Liang, Serena Ferri, Danilo Monarca, Riccardo Massantini. Pine nut species recognition using NIR spectroscopy and image analysis[J]. *Journal of Food Engineering*, 2021, 292(292).
- [9] Hamza Baniata, Ahmad Anaqreh, Attila Kertesz. PF-BTS: A Privacy-Aware Fog-enhanced Blockchain-assisted task scheduling[J]. *Information Processing and Management*, 2021, 58(1).
- [10] Gong Qiuming, Zhou Xiaoxiong, Liu Yongqiang, Han Bei, Yin Lijun. Development of a real-time muck analysis system for assistant intelligence TBM tunnelling[J]. *Tunnelling and Underground Space Technology incorporating Trenchless Technology Research*, 2021, 107(107).
- [11] Zhang Hu, Tang Zhaohui, Xie Yongfang, Gao Xiaoliang, Chen Qing, Gui Weihua. Long short-term memory-based grade monitoring in froth flotation using a

- froth video sequence[J]. Minerals Engineering, 2021, 160(160).
- [12] Wang Mingzhu, Kumar Srinath Shiv, Cheng Jack C.P..Automated sewer pipe defect tracking in CCTV videos based on defect detection and metric learning[J]. Automation in Construction, 2021, 121(121).
- [13] Tuan Nguyen Manh, Hau Quach Van, Chin Sangyoong, Park Seunghee.In-situ concrete slump test incorporating deep learning and stereo vision[J]. Automation in Construction, 2021, 121(121)2.
- [14] Boxin Guan, Yuhai Zhao, Yuan Li. An improved ant colony optimization with an automatic updating mechanism for constraint satisfaction problems[J].Expert Systems With Applications, 2021, 164(prepublish).
- [15] Avalos Edgar, Akagi Kazuto, Nishiura Yasumasa.Visible fingerprint of X-ray images of epoxy resins using singular value decomposition of deep learning features[J]. Computational Materials Science, 2021, 186(186).
- [16] Ananda S. Vaka, Suvankar Ganguly, Prabal Talukdar.Novel inverse heat transfer methodology for estimation of unknown interfacial heat flux of a continuous casting mould: A complete three-dimensional thermal analysis of an industrial slab mould[J]. International Journal of Thermal Sciences, 2021, 160(160).
- [17] Liwei Ju, Liling Huang, Hongyu Lin, Hanfang Li, Zhongfu Tan.An interactive dispatching strategy for micro energy grids considering multi-energy flexible conversion based on the three-level optimization perspective[J]. Sustainable Cities and Society, 2021, 64(prepublish):
- [18] Psota E.T., Luc E.K., Pighetti G.M., Schneider L.G., Trout Fryxell R.T., Keele J.W., Kuehn L.A..Development and validation of a neural network for the automated detection of horn flies on cattle[J]. Computers and Electronics in Agriculture, 2021, 180(180).
- [19] Chen Jiqing, Qiang Hu, Wu Jiahua, Xu Guanwen, Wang Zhikui.Navigation path extraction for greenhouse cucumber-picking robots using the prediction-point Hough transform[J]. Computers and Electronics in Agriculture, 2021, 180(180).
- [20] Wei-Lung Mao, Wei-Chun Chen, Chien-Tsung Wang, Yu-Hao Lin.Recycling waste classification using optimized convolutional neural network[J]. Resources, Conservation & Recycling, 2021, 164(164).

# The Catalytic Process of C4 Olefins Prepared from Ethanol was Investigated based on Multiple linear Regression

Zhenbin Zhang<sup>1\*</sup>, Bingbing Shi<sup>2</sup>, Haoran Bai<sup>1</sup>

<sup>1</sup> College of Science, North China University of Science and Technology, Tangshan 063210, China;

<sup>2</sup> School of Artificial Intelligence, North China University of Technology, Tangshan 063210, China;

\*Corresponding Author.

**Abstract:** C4 olefin is one of the important petrochemical basic raw materials. Ethanol is the raw material for the production and preparation of C4 olefin. At present, there are few specific studies on improving the selectivity of the gas product propylene in this reaction, so it is of great theoretical significance and application value to explore the technological conditions for the preparation of C4 olefins by ethanol catalytic coupling [1]. The results of the same catalytic combination reaction with different charging methods are slightly different. Therefore, multiple linear regression models of four catalyst factors, temperature on ethanol conversion and C4 olefin selectivity are established for different charging methods, and accurate fitting analysis is carried out. In order to enhance the reliability of the results, the final integration model is obtained by using the idea of ensemble learning and adding the Relief F algorithm model of feature extraction.

**Keywords:** Multiple linear regression model; Relief F algorithm integration learning

## 1. PROBLEM ANALYSIS

By analyzing the data, it was found that there were slight differences in the results of the same catalytic combination reaction under different charging methods. Therefore, multiple linear regression models of four factors of catalyst, temperature on ethanol conversion and C4 olefins selectivity were established for different charging methods, and accurate fitting analysis was carried out. In order to enhance the reliability of the results, the final integration model is obtained by using the idea of ensemble learning and adding the Relief F algorithm model of feature extraction. The weights of five factors obtained by the two models were averaged to obtain the effects of different catalyst combinations and temperatures on ethanol conversion and C4 olefins selectivity.

## 2 DEFINITION AND SYMBOL DESCRIPTION

In order to simplify problem analysis and data processing, symbols are stipulated as follows:

symbol	define
$\alpha$	Ethanol conversion
$\beta$	C4 olefins selectivity
$T$	Reaction temperature
$t$	The reaction time
$s_i$	Catalyst combination variable
$\theta$	C4 olefin yield
$n$	The number of particles

$c_1$	Individual learning factors of particles, also known as individual accelerators
$c_2$	The particle's social learning factor, also known as the social acceleration factor
$\omega$	Inertial weight of velocity

## 3. MULTIPLE LINEAR REGRESSION MODEL

Firstly, based on the data analysis, this paper establishes multiple linear regression models of ethanol conversion rate and C4 olefins selectivity in different charging modes from the following five dimensions.

### 3.1 PREPARATION OF MULTIPLE LINEAR REGRESSION MODEL

#### 1. Data standard interval mapping

The units of the five dimensional indicators selected are different, because the conversion rate of ethanol and the selectivity interval of C4 olefin are within the interval of (0, 100). The data are normalized to the interval of [0, 100], and the calculation method is as follows:

$$\text{Normalized data} = \frac{\text{Maximum value of data} - \text{Original data}}{\text{Maximum} - \text{Minimum value}} * 100(1)$$

#### 2. Multiple linear regression model

Breaking down the catalyst portfolio into four dimensions, Co load  $s_1$  /Co/SiO<sub>2</sub> The quality of loading  $s_2$  and HAP The quality of loading  $s_3$  /Ethanol concentration per minute  $s_4$ , With temperature  $T$  Together constitute the influence of ethanol conversion rate, C4 Five dimensions of olefin selectivity:

$$\alpha_{A1} = \omega_1 s_1 + \omega_2 s_2 + \omega_3 s_3 + \omega_4 s_4 + \omega_5 T \quad (2)$$

$$\alpha_{B1} = \omega'_1 s_1 + \omega'_2 s_2 + \omega'_3 s_3 + \omega'_4 s_4 + \omega'_5 T \quad (3)$$

$$\beta_{A1} = \omega''_1 s_1 + \omega''_2 s_2 + \omega''_3 s_3 + \omega''_4 s_4 + \omega''_5 T \quad (4)$$

$$\beta_{B1} = \omega'''_1 s_1 + \omega'''_2 s_2 + \omega'''_3 s_3 + \omega'''_4 s_4 + \omega'''_5 T \quad (5)$$

$\omega_i$  Denotes the regression coefficient of the equation,  $i = 1\backslash 2\backslash 3\backslash 4\backslash 5$ .

Using Matlab statistical toolbox to establish multiple linear regression equation:

$$[b, bint, r, rint, stats] = regress(Y, X, alpha) \quad (6)$$

Where  $b$  is the regression coefficient,  $bint$  is the confidence interval of the regression coefficient,  $r$  is the residual,  $rint$  is the confidence interval of the residual, and  $alpha$  is the significance level. Stats contains four statistics, F test and estimated error variance  $S^2$  The correlation coefficient  $R^2 \backslash P$ . The correlation coefficient  $R^2$  The closer it is to 1, the more significant the regression equation is.  $F > F_{1-\alpha}(k, n - k - 1)$ , Refused to  $H_0$ ,  $F$  The larger is, the more significant the regression equation is. The probability that corresponds to  $F$   $P < \alpha$  ( $\alpha$  Default is 0.05) When refused to  $H_0$ , The model works. The smaller the estimation error variance is, the more

significant the regression equation is.

### 3.2 THE ESTABLISHMENT AND SOLUTION OF MULTIPLE LINEAR REGRESSION MODEL

1\Function of the relationship between different catalyst combinations and temperature on ethanol conversion:

According to the experimental data of loading mode I and II, multiple linear regression was carried out on the data by Matlab, residual graph was drawn and residual value and confidence interval were given.

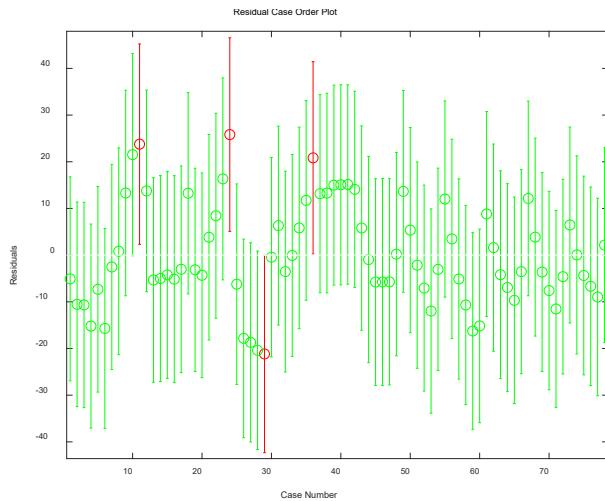


Figure 1. Residual diagram between ethanol conversion rate and factors in Formula I Figure 2. Residual diagram between ethanol conversion rate and factors in Formula II

Table 1 Confidence intervals

Feeding way	The correlation coefficient R <sup>2</sup>	The F value	The probability P that corresponds to F	Estimated error variance
I	0.88716733732681	505.087346125673	0.005264277345	5.821546944089
II	0.8820621886818	12.38438563095022	0.00302113567344	0.9513567897543

I Feeding method: Correlation coefficient R<sup>2</sup> = 0.88716733732681, indicating that the regression equation is very significant. The probability p<α corresponding to F rejects H. According to F test, regression model (7) is established.

$$\alpha_{A1} = 4.8628 - 0.0867s_1 - 0.0010s_2 + 0.2485s_3 + 0.1810s_4 + 0.5545T \quad (7)$$

II Feeding method: Correlation coefficient R<sup>2</sup> = 0.8820621886818, indicating that the regression equation is very significant. The probability p<α

corresponding to F rejects H. According to F test, regression model (8) is established.

$$\alpha_{B1} = -0.9429 + 0.2566s_1 - 0.1714s_2 - 0.1250s_3 + 0.0732s_4 + 0.2933T \quad (8)$$

2\Function of the relationship between different catalyst combinations and temperature selectivity for C4 olefin: In accordance with the experimental data of I and II loading mode, multiple linear regression was carried out on the data by Matlab, residual graph was drawn and residual value and confidence interval were given.

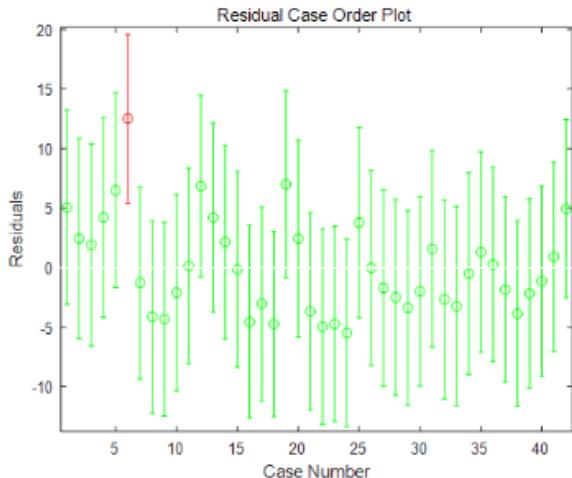


Figure 3. The selectivity of type I C4 alkenes and the residual diagram of factors Figure 4. The selectivity of type II C4 alkenes and the residual diagram of factors

Table 2 Confidence interval II

Feeding way	The correlation coefficient R <sup>2</sup>	The F value	The probability P that corresponds to F	Estimated error variance
I	0.84803693194482	405.087246461155	0.000192859210	5.820027944450
II	0.91880288437719	1404.465658337	0.005345346867	1.325633478064

I Feeding method: Correlation coefficient R<sub>2</sub> = 0.84803693194482, indicating that the regression equation is very significant. The probability p < α corresponding to F rejects H. According to F test, regression model (9) is established.

$$\beta_{A1} = 0.0590s_1 + 0.00005s_2 + 0.00000s_3 - 0.0212s_4 + 0.2730T \quad (9)$$

II Feeding method: Correlation coefficient R<sub>2</sub> = 0.91880288437719, indicating that the regression equation is very significant. The probability p < α corresponding to F rejects H. According to F test, regression model (10) is established.

$$\beta_{B1} = 357.6536 + 0.1238s_1 - 7.3310s_2 + 0.00000s_3 - 0.0389s_4 + 0.4272T \quad (10)$$

#### 4. FEATURE EXTRACTION RELIEF F ALGORITHM MODEL

To strengthen the accuracy of the results, we use the thought of integrated study, set up Relief F algorithm of feature extraction, the calculated results with the results of the multivariate linear regression model superposition analysis, get the final weights of each factor, the analysis under different loading ways, different combination of catalyst and temperature on the influence of ethanol conversion rate, selectivity of C4 olefin.

##### 4.1 PREPARATION AND ESTABLISHMENT OF RELIEF F ALGORITHM MODEL

The calculation steps of the feature extraction algorithm are as follows:

Step1: reset the weights of all factors to 0 and divide them into multiple samples.

Step2: randomly select a sample R, find out the nearest neighbor samples from the heterogeneous and

Table 3 Sample classification of ethanol conversion under loading mode A (part)

$\alpha_A$	$s_2$	$s_1$	$s_3$	$s_4$	T	Classification number
34.05	200	1	200	1.68	250	4
37.43	200	1	200	1.68	275	4
46.94	200	1	200	1.68	300	5
49.70	200	1	200	1.68	325	5
47.21	200	1	200	1.68	350	5
46.24	200	1	200	1.68	400	5
18.07	200	2	200	1.68	250	2
17.28	200	2	200	1.68	275	2
19.60	200	2	200	1.68	300	2
30.62	200	2	200	1.68	325	4
39.10	200	2	200	1.68	350	4
48.40	200	2	200	1.68	400	5
5.50	200	1	200	0.9	250	1
8.04	200	1	200	0.9	275	1
17.01	200	1	200	0.9	300	2
28.72	200	1	200	0.9	325	3
36.85	200	1	200	0.9	350	4
53.43	200	1	200	0.9	400	6
9.62	200	0.5	200	1.68	250	1

#### 4.3 THE RELIEF F ALGORITHM MODEL WAS SOLVED

After data classification was completed, Matlab software was used to solve the problem. According to ReliefF algorithm, the weight of each factor of different catalyst combinations and temperatures on ethanol conversion rate and C4 olefins selectivity in the two charging methods was calculated. The 20 times of weight iterative calculation figure and the average weight table are as follows:

Table 4 Average weight of 20 iterations

homogeneous samples, find the differences in different features of the samples, and obtain the weight of features according to sample analysis.

Step3: repeat the selection operation of step2, obtain samples and feature differences according to the sorted classes, and update the weight W according to the formula.

Step4: average the result of weight W of each cycle.

Sample weight W is calculated as follows:

$$W(A) = W(A) - \sum_{j=1}^k \text{diff}(A, R, H_j) / (mk) + \sum_{c=\text{clas}(R)} \left[ \frac{p(C)}{1-p(\text{Class}(R))} \sum_{j=1}^k \text{diff}(A, R, M_j(C)) \right] / (mk) \quad (11)$$

In the above formula,  $\text{diff}(A, R_1, R_2)$  represents the difference between samples  $R_1$  and  $R_2$  on feature A, and  $M_j(C)$  denotes the nearest sample  $j$  of the  $C$  class sample, as follows:

$$\text{diff}(A, R_1, R_2) = \begin{cases} 0 & R_1[A] = R_2[A] \\ 1 & R_1[A] \neq R_2[A] \end{cases} \quad (12)$$

#### 4.2 DATA CLASSIFICATION

According to the analysis of data and algorithm, the data should be classified first. After comprehensive analysis of different charging methods, we found that the conversion rate of ethanol ranged from 0% to 90%, and the selectivity of C4 olefin ranged from 0% to 60%. Therefore, 10% interval was taken as the span of classification, and 0% to 10% samples were recorded as the first type samples. The sample data of different charging methods are classified according to the above method. Part of the results are as follows, and all the results are shown in the appendix:

Table 3 Sample classification of ethanol conversion under loading mode A (part)

	$s_1$	$s_2$	$s_3$	$s_4$	T
$\alpha_{A2}$	0.0917	0.0020	0.2523	0.2025	0.5238
$\alpha_{B2}$	0.1254	0.5912	0.0051	0.0061	0.4122
$\beta_{A2}$	0.2543	0.1653	0.1432	0.0489	0.2894
$\beta_{B2}$	0.0534	0.0212	0	0.0367	0.2948

#### 5. INTEGRATED ANALYSIS OF MULTIPLE LINEAR REGRESSION MODEL AND RELIEF F ALGORITHM MODEL

The absolute value of the function coefficient obtained by establishing the multiple linear regression model is the weight of each factor, and the symbol represents the

correlation. The absolute value of all coefficients of the multiple linear regression model is expressed in the same way as above, and the results obtained by the Relief F algorithm are written as follows.

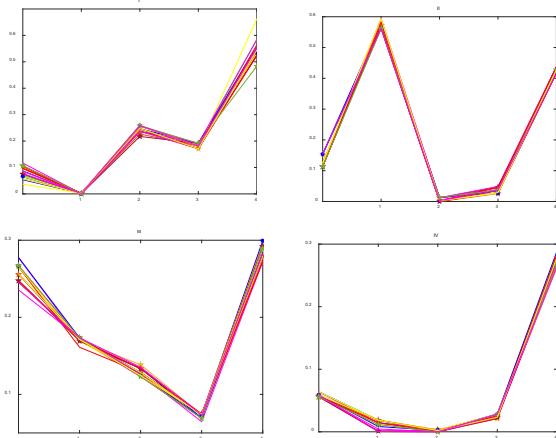


Figure 5. Numerical diagram of weight for 20 iterations  
Table 5 Weight results of various factors under different loading modes in multiple linear regression model

	$s_1$	$s_2$	$s_3$	$s_4$	$T$
$\alpha_{A1}$	0.0867	0.0010	0.2483	0.1810	0.5545
$\alpha_{B1}$	0.1238	7.3310	0	0.0389	0.4272
$\beta_{A1}$	0.2566	0.1714	0.1250	0.0732	0.2933
$\beta_{B1}$	0.0590	0.00005	0	0.0212	0.2730

Table 6 Average weight results of all factors under different loading methods of Relief F algorithm

	$s_1$	$s_2$	$s_3$	$s_4$	$T$
$\alpha_{A2}$	0.0917	0.0020	0.2523	0.2025	0.5238
$\alpha_{B2}$	0.1254	0.5912	0.0051	0.0061	0.4122
$\beta_{A2}$	0.2543	0.1653	0.1432	0.0489	0.2894
$\beta_{B2}$	0.0534	0.0212	0	0.0367	0.2948

According to the above two tables, the analysis results of influences of various factors on ethanol conversion rate under different bagging methods are as follows:

Table 7 Analysis of factors and degree that affect ethanol conversion rate

Loading way	Factors affecting ethanol conversion rate	Degree of influence	The weight
I	The temperature	high	0.55
	HAP mass, rate of ethanol infusion	Medium	0.25\0.20
II	The quality of the Co	high	0.59
	The temperature	Medium	0.41

Table 8 Factors and degree analysis of influencing C4 olefins selectivity

Loading way	Factors affecting ethanol conversion rate	Degree of influence	The weight
I	Temperature, Co/SiO2 quality	high	0.28\0.25
	Co quality, HAP quality	high	0.16\0.14
II	The temperature	high	0.26

Co/SiO2 quality\		
The rate of ethanol infusion	low	0.05\0.03

1. In I charging mode, temperature factor has a great influence on ethanol conversion, with a weight of about 0.55; The influences of HAP mass and ethanol infusion rate on ethanol conversion were the next, with weights of 0.25 and 0.20.

2. Under the loading mode II, the weight ratio of Co load in the linear regression was 7.33, which was significantly different from the results calculated by Relief F algorithm, but compared with other factors, Co load still had a great impact on ethanol conversion rate; The influence of temperature on ethanol conversion was the second, and the weight was about 0.41.

Similarly, the analysis results of the influence of various factors on the selectivity of C4 olefin in different bagging methods are as follows:

3. In I charging mode, temperature and Co/SiO2 mass have great influence on the selectivity of C4 olefins, with weights of about 0.28 and 0.25. Co loading and HAP mass had the second most significant effect on C4 olefins selectivity, with weights of 0.16 and 0.14.

4. In II charging mode, temperature has a great influence on the selectivity of C4 olefin, and the weight is about 0.29. The quality of Co/SiO2 and the rate of ethanol infusion had the second effect on the selectivity of C4 olefins, with weights of about 0.05 and 0.03.

#### 6. Evaluation of the model

Relief series of algorithms have high operation efficiency and no restriction on data types. They belong to a feature weight algorithm, which gives high weight to all features with high correlation with categories. Therefore, the limitation of the algorithm is that it cannot effectively remove redundant features.

#### REFERENCES

- [1] LU SHAopei. Synthesis of butanol and C4 olefin by ethanol coupling [D]. Dalian University of Technology, 2018.
- [2] LIX L. Research and application of Feature selection algorithm based on Relief [D]. Dalian University of Technology, 2013.
- [3] Xu Jiwei, Yang Yun. Journal of yunnan university (natural science), 2018, 40 (06):1082-1092.
- [4] Dong Qiaoling. Research on application of total least square method in multiple linear regression under different error influence model [D]. Taiyuan University of Technology, 2016.

# Research on Converter Flue Gas Analysis Based on Penalty Factor Improved Machine Learning

Yi Zhu<sup>1, 2</sup>, Yibo Wang<sup>1, 2</sup>, Yunjiang Han<sup>3</sup>

<sup>1</sup> Yi Sheng College, North China University of Science and Technology, Tangshan 063210, China;

<sup>2</sup> Mathematical Modeling Innovation Lab, North China University of Science and Technology, Tangshan 063210, China;

<sup>3</sup> College of Mechanical and Electrical Engineering, Shenyang Aerospace University, Shenyang 110135, China

**Abstract:** In order to realize intelligent steelmaking and improve the end-point hit rate, a random forest model and a BP neural network model are established to analyze and predict the carbon content and temperature in molten steel. In order to make the model always have a good prediction effect between 70%-85%, the penalty factor is introduced as an independent variable, the error term obtained after the original model predicts the data is multiplied by the penalty coefficient  $\lambda$  to obtain the penalty factor, and the model Where  $\lambda$  is 1. The new variables after adding the penalty factor are retrained in the optimized model, and an improved model with more accurate prediction effect in the interval of [70%, 85%] is obtained. After testing, the improved random forest model is used to predict the carbon content of molten steel in [70%, 75%] this fixed interval always has more accurate results; the improved BP neural network model is used to predict the temperature of molten steel in [70%, 85%] There are always more accurate results in this fixed interval.

**Keywords:** Random Forest; BP Neural Network; Penalty Factor; Converter Flue Gas

## 1. INTRODUCTION

Smart steelmaking is a necessary way for small and medium-sized enterprises in my country's iron and steel industry to face the coupling pressure of "de-capacity, green manufacturing, and market competition" to realize the transformation of enterprise technology structure and complete enterprise development. The key to intelligent steelmaking is to accurately control the carbon content and temperature of molten steel. Using intelligent algorithms, a mathematical model can be established between the carbon content and temperature of molten steel and the main factors affecting the two to achieve the Effective prediction of carbon content in water and molten steel temperature<sup>[1-2]</sup>.

## 2 [C]&[T] PREDICTION MODEL BASED ON MACHINE LEARNING

### 2.1 DATA SOURCES

Based on an experiment, the proportion of oxygen consumption, total oxygen, CO content in flue gas, CO<sub>2</sub> content in flue gas and carbon content in molten steel [C] and molten steel temperature value [T], which contains 5 data sets, analyzes the correlation between the indicators, takes the first four variables as independent variables, and the last two variables as dependent variables.

### 2.2 [C]&[T] PREDICTION MODEL BASED ON BP NEURAL NETWORK

The general structure of a neural network is composed of

an input layer, a hidden layer, and an output layer. The hidden layer can be one layer or multiple layers, and the layers are connected to each other. The loss is obtained by forward propagation, and the error is returned by BP, and the weight of each layer is corrected according to the error signal. Differentiate each  $\omega$ , and then update each  $\omega$ .

**The number of hidden layers is determined:** An increase in the number of network layers will improve the accuracy and speed of operations<sup>[3]</sup>. However, the complexity of the model has also increased, and the risk of overfitting has also increased. Considering the small number of data samples, a 10-layer network is selected to build the model.

**Selection of activation function:** Since the derivative values of the Sigmoid function and the Tanh function are not greater than 1, then the gradient will inevitably disappear during the iterative process. The ReLU function can avoid this phenomenon well, so this paper chooses the ReLU function as the activation function of the model in this paper.

**Selection of training algorithm:** In the process of neural network training, three algorithms, Levenberg-Marquardt, Bayesian regularization, and quantized conjugate gradient are generally used for training<sup>[4]</sup>. Since the data samples used in this article are small, over-fitting is not easy to appear, and it will not take too much training time, so the Bayesian regularization training method is used. Based on the above analysis, the following neural network structure is established:

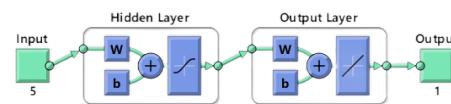


Figure 1 Neural network structure

### 2.3 [C]&[T] PREDICTION MODEL BASED ON RANDOM FOREST

Random forest is a special bagging method that uses decision trees as a model in bagging<sup>[5]</sup>. First, use the bootstrap method to generate m training sets. Then, for each training set, construct a decision tree. When the node finds the features for splitting, randomly extract a part of the features from the features, and find the best among the extracted features<sup>[6-8]</sup>. The optimal solution is applied to the node for splitting. Assuming that X and Y are input and output variables, respectively, and Y is a continuous variable, given the training data set:

$$D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$$

It can be defined as different regions by traversing each feature variable and its corresponding feature value of the

training set data. Assuming that the segmentation variable being divided into regions is the  $i$ -th variable, and the corresponding segmentation feature value is  $j$ , then:

$$\begin{cases} a_1(i, j) = \{x | x^{(i)} \leq j\} \\ a_2(i, j) = \{x | x^{(i)} > j\} \end{cases}$$

Through continuous traversal and division, the input data is divided into  $L$  subspaces  $a_1, a_2, \dots, a_L$ , and each subspace  $a_1$  contains its corresponding sample data and output value  $\beta_1$ , then the solution of the model at this time is :

$$f(x) = \sum_{l=1}^L \beta_l I(x \in a_l)$$

Using the sum variance measurement, the measurement goal is to obtain the minimum mean square error of each set of  $D_1$  and  $D_2$  for the data sets  $D_1$  and  $D_2$  on both sides of the partition point  $j$  for the division feature  $i$ , and the minimum sum of the mean square errors of  $D_1$  and  $D_2$ . The expression is:

$$\min_{i,j} \left[ \min_{\beta_1} \sum_{x_i \in a_1(i,j)} (y_i - \beta_1)^2 + \min_{\beta_2} \sum_{x_i \in a_2(i,j)} (y_i - \beta_2)^2 \right]$$

#### Determination of the number of trees and the number of leaf nodes:

As can be seen from the figure below, in the prediction model of [C] and [T], when the number of decision trees and leaf nodes are 5 and 20, respectively, the random forest model exhibits the most ideal mean square error;

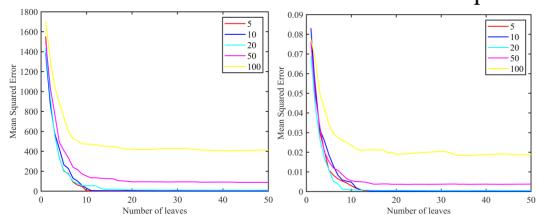


Figure 2 Adjustment of the number of leaves in random forest

In the same way, draw the curve formed by the trees in the random forest model and their mean square error. The figure below shows that when predicting [C] and [T], the number of trees is both 20. When the model error converges to the minimum, this paper determines the number of trees in the random forest as 20.

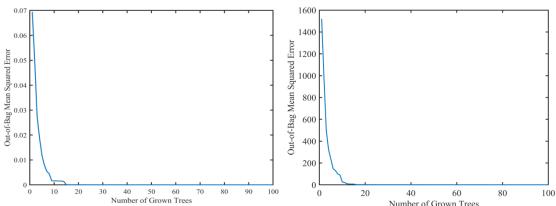


Figure 3 Adjusting the number of trees in a random forest  
2.4 ANALYSIS OF MODEL RESULTS

Based on the above analysis, the neural network models of [C] and [T] are established respectively, and the model training error is shown in the following figure:

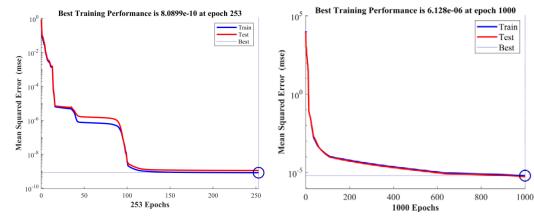


Figure 4 Neural network training effect

The figure reflects that [C] neural network model training error and test error gradually converge after about 100 iterations; [T] neural network model training error and test error gradually converge after about 600 iterations, and the final errors are all in Within 105, it shows that the training effect is better.

After establishing a random forest model for the data sets of [C] and [T], calculate the goodness of fit, and the results are shown in the figure below. It can be seen from the figure that the fitting effect of [C] and [T] based on the random forest model is better, and the  $R^2$  of both are close to 1, indicating that the model establishment effect is better.

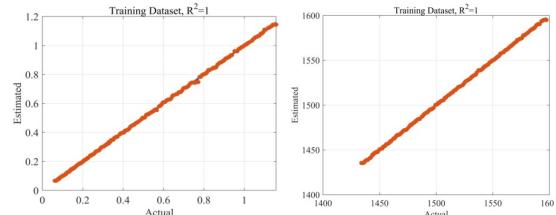


Figure 5 Fitting effect of random forest

Comparing the errors obtained from the BP neural network and the random forest test set, the analysis of the following figure shows that only when the two are used for the fitting prediction of their own data set, the errors of the BP neural network in the [C] model and the [T] model are both It is much smaller than the error of random forest to data prediction.

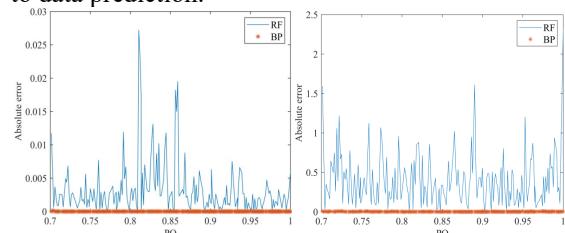


Figure 6 Comparison of the prediction effects of the two models used in their own data sets

#### 3. IMPROVEMENT OF MACHINE LEARNING MODEL BASED ON ADDITIONAL PENALTY ITEMS

The penalty term method is a model structure optimization algorithm that improves its generalization ability by indirectly pruning the network structure. The principle is to add a "complexity item" to the traditional error function to measure the complexity of the network structure and reflect the pros and cons of generalization ability. This item inhibits the increase of the weight value during the training process and drives the unimportant weight value to gradually decrease. Small to zero effect, so that the model adapts to the known samples while ensuring that its structure is not too complicated [9]. Assuming that that the vector made by the ownership parameters in the model is

$w$ , then the error function with penalty term can be described as follows:

$$E(w) = E(w) + \lambda E_p(w)$$

The first term at the right end of the above formula is the standard performance measurement of the traditional error function on the network. It depends on the network structure model and input data at the same time. It is generally defined as the mean square error or square error on the training sample set. The second term  $\lambda E_p(w)$  is a penalty term, which only depends on the network structure model parameters. It is called the penalty term parameter or penalty term coefficient, which represents the relative importance of the penalty term with respect to the performance measurement item.

It can be seen from the above error image that whether it is an optimized random forest model or a BP neural network model, there is not much difference in the error generated when predicting PQ between 70%-85% and 85%-100%. Make the prediction of the two models more accurate when the PQ is between 70% and 85%. Now based on the weighted average model, after introducing the penalty factor, the weighted random forest model and the BP neural network model are established again to achieve the PQ in the target of higher model accuracy between 70%-85%.

### 3.1 THE ESTABLISHMENT OF A MACHINE LEARNING MODEL WITH THE INTRODUCTION OF PENALTY FACTORS

Suppose the penalty factor is  $\lambda L_i$ , where  $\lambda$  is the penalty coefficient, which is set to 1 in this article. In order to make the prediction result more accurate when the model is between 70%-85%, the penalty factor is now defined as follows:

$$\lambda L_i = \begin{cases} A_i - B_i, & 70\% \leq PQ_i \leq 85\% \\ 0, & 85\% \leq PQ_i \leq 100\% \end{cases}$$

When PQ is in the range of 70%-85%, the error between the predicted value of the original model and the true value is set as the penalty factor; when PQ is in the range of 85%-100%, the value of the penalty factor is set to 0. As a result, this column of data is set as a new variable, and back to the model for training.

### 3.2 IMPROVED ANALYSIS OF MODEL RESULTS

After the training of the two models after introducing the penalty factor is completed, the other data sets are respectively predicted, and the result is subtracted from the true value to obtain the absolute error, and the accuracy of the PQ in different intervals is calculated after inversion.:

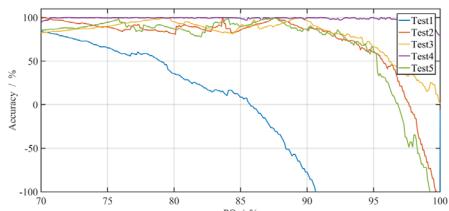


Figure 7 The accuracy of the random forest model improved based on the penalty factor

From the analysis of the above figure, after adding the penalty term, the RF model has a better prediction effect on the data when the PQ is between 75% and 80% than when the PQ is between 85% and 100%. When predicting data set 1, the accuracy reaches 60% when PQ is between 75% and 80%, and as the PQ index increases, the model's prediction accuracy for data set 1 decreases. And when the PQ reaches 85%, the prediction accuracy of the random forest model for data set 1 is reduced to 0%, indicating that the random forest model with the penalty term is not good for the prediction of data set 1; However, when the improved model predicts data sets 2, 3, 4, and 5, the accuracy is above 85% when the PQ is between 75% and 80%, and the accuracy of the model only drops when the PQ is between 95% and 100%. Which reflects that the improved model has a good prediction effect for data sets 2, 3, 4, and 5 when the PQ is between 75% and 95%. In summary, the improved random forest model always has a better prediction effect in the fixed interval of 70% - 75%.

In the same way, after the neural network model with the penalty factor introduced is trained, predictions are made on other data sets respectively, and their accuracy is calculated as shown in the figure below:

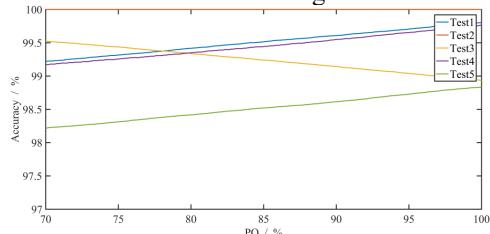


Figure 8 The accuracy of the improved neural network model based on the penalty factor

From the graph analysis, it can be seen that, compared with the random forest model, the improved neural network has a generally better prediction effect for [T], and its accuracy is above 98%, and the accuracy is linear with the change of PQ. For data sets 1, 4, and 5, the prediction effect of the neural network increases with the increase of PQ; for data set 3, the accuracy of the neural network's prediction decreases with the increase of PQ. Since the improved neural network has a higher accuracy for the data set 1-5 in the data set, the fixed interval range is determined to be 70%-85%.

### 4. CONCLUSION

Based on the above analysis, the improved random forest and neural network model has a better prediction effect in the specified interval of most data sets. The accuracy in this interval is due to the outside of the interval, indicating that the addition of the penalty factor improves the model's prediction of the specified interval Effect.

For [C] predicted by the improved random forest model, the accuracy improvement effect of the model for different data sets is different: for data set 1, the accuracy improvement is more obvious when the PQ is 81%-85%, while for the data In sets 2, 3, the improved random forest prediction model has a significant improvement in accuracy when the PQ is between 87%-90% and 90%-83%. On the whole, the improved model has a slight

increase in overall accuracy compared to the original model. For [T] predicted by the improved BP neural network model, there are still big differences in the accuracy of the model for different data sets: The improved model's accuracy improvement for data set 1 is mainly concentrated in the larger range of PQ; the accuracy of data set 4 is also relatively large, but there is no obvious range; the accuracy of data set 2, 3 is slightly improved.

In summary, on the whole, the improved model has a certain increase in prediction accuracy compared to the original model, but the impact on the accuracy of different ranges of PQ for different data sets is different. Compared with the original model, the deep learning model after weighted average processing and adding penalty factors has better accuracy and robustness.

#### REFERENCES

- [1] Zhou Hang, Wang Xiao, Li Peng, Deng Yuping.Optimization of converter automatic steelmaking process model based on flue gas analysis[J]. Hebei Metallurgy, 2020(10):48-51+82.
- [2] Sun Jiangbo. Converter dynamic control model based on flue gas analysis[D]. Hebei Union University, 2014
- [3] Zhao Guangyuan, Ma Fei.Dust concentration prediction of BP neural network optimized by particle swarm optimization[J]. Measurement and Control Technology, 2018, 37(06):20-23.
- [4] Du Liang, Research on the optimization of selected laser melting process based on neural network and genetic algorithm[D]. Xiamen University of Technology, 2021.
- [5] Wang Hong, Zhang Qiang, Wang Ying, Guo Yujie.Improved CART decision tree regression algorithm based on ELM[J]. Computer Systems Applications, 2021, 30(02):201-206.
- [6] Zheng Xuezao, Li Menghan, Zhang Yinni, Jiang Peng, Wang Baoyuan.Study on prediction model of coal spontaneous combustion temperature based on random forest algorithm[J]. Industry and Mine Automation, 2021, 47(05):58-64.
- [7] Xia Huiwei. Research on network traffic analysis and prediction based on decision tree integration and wide forest[D]. Nanjing University of Posts and Telecommunications, 2020.
- [8] Dong Shishi, Huang philosophy. Analysis of Random Forest Theory[J]. Integration Technology, 2013, 2(01):1-7.
- [9] Lu Qiongyu, Tong Xuefeng.Research on the improvement of BP algorithm[J]. Computer Engineering and Design, 2007(03):648-650.
- [10] Shao Hongmei. Convergence of BP neural network training algorithm with penalty term[D]. Dalian University of Technology, 2006.