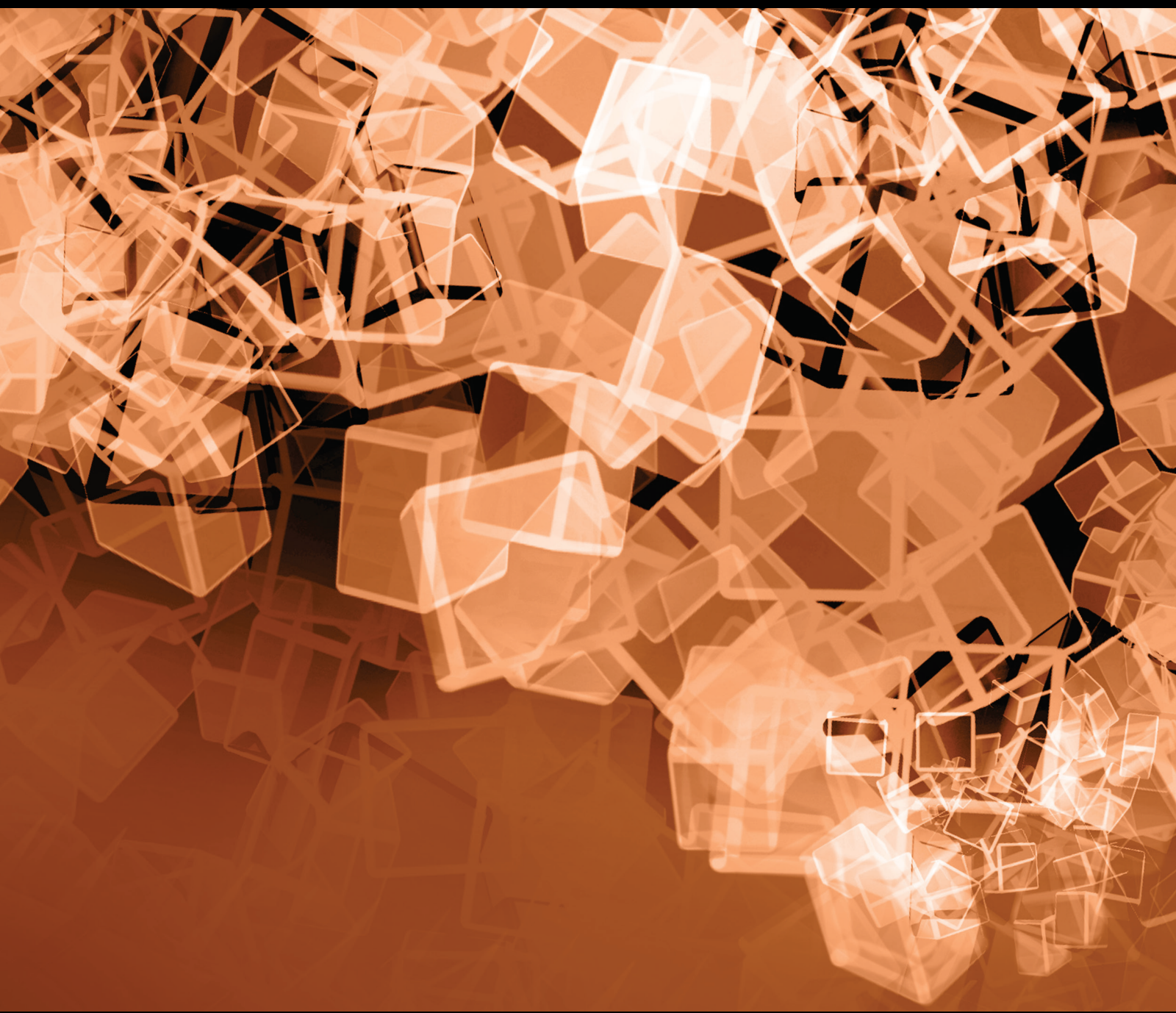
Discrete Dynamics in Nature and Society

Multi-Goal Decision Making for Applications in Nature and Society

Lead Guest Editor: Chi-Hua Chen
Guest Editors: Xianbiao Hu, Guo Wei, and Fuqiang Gu



Multi-Goal Decision Making for Applications in Nature and Society

Discrete Dynamics in Nature and Society

Multi-Goal Decision Making for Applications in Nature and Society

Lead Guest Editor: Chi-Hua Chen
Guest Editors: Xianbiao Hu, Guo Wei, and Fuqiang Gu

Contents

Contents

Hindawi
Discrete Dynamics in Nature and Society
Volume 2020, Article ID 6408654, 11 pages
https://doi.org/10.1155/2020/6408654

Research Article

The Impact of R&D Input on Technological Innovation: Evidence from South Asian and Southeast Asian Countries

Lei Lv, Yuchen Yin, and Yuanchang Wang

School of Mathematics, Yunnan Normal University, Kunming 650500, China

Correspondence should be addressed to Yuanchang Wang; wangyuanchang@ynnu.edu.cn

Received 25 May 2020; Revised 15 July 2020; Accepted 13 September 2020; Published 7 December 2020

Academic Editor: Chi-Hua Chen

In the era of the rapid development of knowledge economy and science, all countries have thought highly of technical innovation and greatly increased the R&D input for it. However, the research on the impact of R&D input on technical innovation lacks specialized, cross-country, and cross-time investigations, and especially, the research on small countries such as South Asia and Southeast Asia where technical innovation is relatively backward. So, does R&D input in South Asia and Southeast Asia have an impact on technical innovation and to what extent? Let us analyze the panel data of 18 countries in South Asia and Southeast Asia from 2001 to 2018, use three methods of unit root test to test the stationarity of variables, adopt the Kao cointegration test to test a stable long-term relationship between the variables, and then, respectively, carry out the transnational regression analysis of the difference between patent applications, scientific journal articles, and the R&D input with multiple models. Finally, the heteroscedasticity robust fixed-effect model is found to be the most suitable for this study after the comparative analysis of multiple models. Through the fixed-effect intercepts of each country in the heteroscedasticity robust fixed-effect models, South Asian and Southeast Asian countries are divided into three levels, and each level lists a set of equations. So, the following conclusions are drawn: both R&D expenditure and manpower input in South Asia and Southeast Asia significantly promote technological innovation; the efficiency of both R&D expenditure and manpower input promoting technological innovation in South Asia and Southeast Asia is low and needs to be improved. These conclusions provide references for policymakers in some countries where technological innovation is relatively backward, especially in Cambodia, Lao PDR, Myanmar, Timor-Leste, Bangladesh, Bhutan, Maldives, and Nepal.

1. Introduction

In this time of the high-speed development of information technology and knowledge economy, technology innovation has increasingly become an engine of social development, a decisive factor for all countries to enhance their international competitiveness and the vital motive of the continuous economic growth of a country or a region. Both developed and developing countries are increasingly aware of the importance of technological innovation in various aspects of economic and social development, even some small countries in South Asia and Southeast Asia where innovation is relatively backward are no exception. Improving the level of technology input has become the only way for countries to enhance their national comprehensive strength. Moreover, research and development (R&D) is a major indicator to internationally measure the level of technology input, the innovation capacity, sustainable development potential, comprehensive technology strength, and competitiveness of a country or region in the world. Therefore, let us select the panel data of 18 countries in South Asia and Southeast Asia from 2001 to 2018 as the research sample to study the impact of R&D input on technology innovation, which is not only conducive to promoting the improvement of the technological level of South Asian and Southeast Asian countries and enhancing the comprehensive strength of each country, but also has great theoretical value and practical significance.

Griliches developed the framework of knowledge production functions, and its framework could be simply expressed as $R\&D\ output = F(R\&D\ input)$. Meanwhile, the production function of C-D type was expressed as

$R\&D\ output = c(R\&D\ input)^{\beta}$, where c represented a constant term and β denoted the output elasticity of R&D input [1]. Subsequently, Jaffe extended the framework of Griliches' knowledge production function, integrating the new economic knowledge pursued by enterprises into the production process, and the input variables included expenditure and manpower input for R&D [2]. Romer, a representative of endogenous growth theory, proposed that knowledge accumulation and knowledge spillover effect of production capacity could make knowledge the main cause of technological innovation and economic growth [3]. After that, Jones revised Romer's knowledge production function and improved the theory and summarized Romer–Jones knowledge production function as $\widehat{A} = \delta L^{\lambda} A^{\phi}$. Among them, the term $\widehat{A}$ represented the knowledge production, L denoted the R&D manpower input, A stood for the stock of knowledge, and δ was the synthesis of other external factors affecting knowledge production. The function showed that the production of new knowledge was the result of the joint action of the R&D manpower input and the external factors of knowledge production and the knowledge stock. Obviously, R&D expenditure investment belonged to the category of external influencing factor. According to this model, the scale of R&D input was a decisive factor for the production of new knowledge when the knowledge stock and the R&D manpower input were relatively constant [4].

In view of the relationship between R&D input and technology innovation, most scholars believed that R&D input had a positive impact on innovation, while a few scholars held the opposite attitude. Scheele, the first to make a quantitative study on the two, used the number of patents as an indicator to evaluate innovation, analyzed the data of 448 large American enterprises by establishing a linear model of R&D input, sales, monopoly, and innovation, and empirically obtained that R&D input had a positive impact on innovation performance [5]. Subsequently, Pakes and Griliches analyzed the data of 448 large American enterprises and found a significant positive relationship between patent applications and R&D input [6]. Crepon et al. believed that R&D activities were divided into two processes, namely, input and output. Among them, R&D expenditure was selected as input factor, and sales of patents and new products were selected as output factors to measure the relationship between input and output, and then the results showed that there was a positive relationship between the two [7]. Nasierowski and Arcelus obtained a positive correlation between R&D innovation efficiency and R&D production capacity through the DEA method [8]. Pellegrino et al. classified R&D input into two categories, specifically R&D expenditures and personnel input, which were positively correlated with patent output [9]. Bogliacino and Pianta believed that R&D input intensity had a significant role in boosting the innovation output of the industry and also played a positive role in promoting profit growth [10]. Baumann and Kritikos believed that the benefits from innovation in microenterprises were basically similar to those in large companies and the boost brought by R&D input intensity to innovation in microenterprises was not much different from that in large companies [11]. Based on the analysis of Chinese regional innovation data in the past five years, Zhu et al. empirically tested that R&D input generally had a significant positive impact on regional innovation performance under the premise of infrastructure control [12]. Moreover, Jefferson et al. and Le et al. also believed that R&D input and innovation output were positively correlated [13, 14].

However, a few scholars had concluded that R&D input did not always improve enterprise performance and uncertainty was an inherent characteristic of R&D activities, which could inhibit the promotion of R&D activities to innovation output [15, 16]. Compared with the high risk of returns, the probability of success of innovation activities was also relatively low, and the risk losses are likely to exceed returns. In the late 20th century, MOTOROLA invested in the iridium R&D project, which ended in failure and cost the company $5 billion [17]. This case became a typical representative of the failure of R&D activities. Hitt et al. believed that the input was negatively correlated with enterprise performance because the risk of R&D input was too high, so the enterprises with good business performance had little risk and were unwilling to take risks in R&D activities [18]. Shi confirmed through research that R&D input would increase the risk and premium level of bonds, worsen the bond rating, damage the interests of creditors, and reduce the refinancing opportunities of the company [19]. In order to further verify the relationship between the two, Bin and Koellinger had obtained similar results by collecting data from different industries [20, 21]. For the manufacturing and information technology industry, Lu and Wang used the data of 99 listed companies in the manufacturing and information technology industry from 2005 to 2008 to empirically analyze that R&D input harms corporate performance [22]. For the traditional manufacturing industry, Qu and Lu adopted the parametric stochastic frontier analysis method to study and found that the internal expenditures of R&D funds had a certain negative impact on the traditional manufacturing industry due to the unreasonable structure [23]. In a word, there are few studies on the negative inhibition effect.

To sum up, according to the theory of knowledge production function, most of the empirical research literature had illustrated the impact of R&D input on technological innovation from various aspects, which had both positive promotion relationships and negative inhibition impact. But there are relatively fewer studies that have negative inhibitory effects. So there were two key points of disagreement among existing research conclusions. One was that scholars had different opinions on index selection, evaluation method, function mode, and performance size in empirical research. Secondly, in different periods, different industries, and different countries or regions, the impact of R&D input on technological innovation showed different characteristics. So, there was still a gap in existing studies, namely, the lack of cross-border and cross-time studies on the macro data at the national level, especially the studies of some small countries in South Asia and Southeast Asia. Furthermore, does R&D input in these small countries have an impact on technological innovation and to what extent? Let us analyze the panel data of 18 countries in South and Southeast Asia from 2001 to 2018, estimate multiple models and select the

best model for the relationship between R&D investment and technological innovation, so as to supplement the existing research, provide basic research for other countries to deepen technology cooperation with South and Southeast Asian countries, provide some enlightenment for the analysis of relevant technology policies, and provide a reference for policymakers in some countries where technological innovation is relatively backward, especially in Cambodia, Lao PDR, Myanmar, Timor-Leste, Bangladesh, Bhutan, Maldives, and Nepal.

2. Model Design

2.1. Variables and Data. Patent applications and scientific journal articles are regarded as dependent variables in this paper. Although the academic community has not reached a consensus on the quantitative indicators of technology innovation, most scholars measure the innovation outputs from the perspectives of patents and scientific journal articles, such as Chen et al. and Wang and Huang [24, 25]. Patents are particularly important because they are a kind of means for innovators to exchange knowledge and create economic value in the market [26]. So, Fritsoh and Franlce and Buesa et al. had emphasized that patents were a useful quantitative indicator of technology innovation [27, 28]. In fact, not all innovations are submitted as patents or other forms of technological achievements, and there are many other ways to transform knowledge into market value [29, 30]. The number of scientific journal articles is one of the main indicators to measure the performance of knowledge creation and is also the main channel for research institutions and universities to provide local actors with technical outputs because it reflects the quality of technological activities in the region [31, 32].

R&D expenditure investment and R&D manpower input are taken as independent variables in this paper. R&D expenditure investment is the primary factor influencing technological innovation and transformation, reflecting the R&D scale of a region, as well as the basic innovation capacity and potential technological output capacity of a region [33, 34]. So, Yoo had also emphasized that R&D expenditure investment was a useful quantitative indicator of R&D input [35, 36]. However, the use of a single variable to measure R&D input is flawed, and it is easy to affect the empirical results due to the error of variable selection [37]. Moreover, the R&D manpower input also represents the R&D scale of the region and also reflects the technical innovation capacity and output capacity [38]. Higher technological innovation ability can also promote R&D expenditure investment to a certain extent [39]. Therefore, more existing studies have quantified the level of R&D input through the R&D expenditure and manpower input [40].

The control variables are selected from the following three aspects in this paper. Firstly, foreign direct investment partly reflects the economic openness level of a country [41, 42]. The level of economic openness has no direct impact on R&D input, but it can increase the technological output of a region by improving the efficiency of R&D input and independent innovation capabilities [43, 44]. Secondly, government education expenditure provides the accumulation of knowledge for national scientific development and innovation and trains the talents needed [45, 46]. Thirdly, regional innovation ability is affected by the regional economic development level to a certain extent, so the imbalance of each regional economic development level also determines the regional differences in technological innovation [47]. And GDP per capita is an important indicator of the macroeconomic performance of a country or region [48]. In short, it is necessary to regard the above three aspects as control variables in this paper. Table 1 shows the settings and descriptions of all the above variables.

The data in this paper are the panel data of all countries in South Asia and Southeast Asia from 2001 to 2018. The main data source is the relevant statistics released on *Global Innovation Index 2018* and the official website of the World Bank. With the help of Stata 15.1 and EViews 8 analysis software, the estimation and analysis of multiple models including the mixed least square model, the random-effect model, the fixed-effect model, and the heteroscedasticity robust fixed-effect model are carried out. The missing data is obtained by trend extrapolation and multiple imputations and all variables adopt the winsorization method to conduct 1% double-tail reduction processing to eliminate the influence of extreme values on modeling analysis. As the cycle of technology innovation is generally long, the time is divided into six stages (2001–2003, 2004–2006, 2007–2009, 2010–2012, 2013–2015, and 2016–2018). Correspondingly, the time is 1–6. The variables at each stage are averaged.

2.2. Theoretical Analysis and Basic Models. C-D knowledge production function is a vital theoretical model to study knowledge production, technological innovation, and its determinants [49]. In order to get the relationship between R&D input and technological innovation in South and Southeast Asian countries, we introduce R&D expenditure investment and manpower input into a typical C-D knowledge production function:

$$\begin{aligned} sja_{it} &= Arei_{it}^{\alpha_1} rmi_{it}^{\alpha_2}, \\ pat_{jt} &= Brei_{jt}^{\beta_1} rmi_{jt}^{\beta_2}, \end{aligned} \tag{1}$$

where α_1, β_1 and α_2, β_2 are the output elasticity of R&D expenditure investment and manpower input, respectively. In addition, A and B are both technological innovation efficiency, which are usually constant terms.

For reducing heteroscedasticity and increasing explanatory power of results, all variables are processed with log transformation to obtain the following model:

$$lnsja_{it} = lnA + \alpha_1 \times lnrei_{it} + \alpha_2 \times lnrmi_{it} + \omega_{it}, \tag{2}$$

$$lnpat_{jt} = lnB + \beta_1 \times lnrei_{jt} + \beta_2 \times lnrmi_{jt} + \varphi_{jt}, \tag{3}$$

where lnsja means logarithm of the scientific journal articles, lnpat denotes logarithm of the patent applications, lnrei means logarithm of the R&D expenditure investment, lnrmi denotes logarithm of the R&D manpower input, and ω_{it} and ϕ_{jt} are both random error terms.

Table 1: Variable settings and descriptions.

Variable name	Variable index	Description
Scientific journal articles	sja	The number of scientific and engineering articles published in the fields of physics, biology, chemistry, mathematics, clinical medicine, biomedical research, engineering and technology, Earth and space science, etc.
Patent applications	pat	The number of patent applications filed through the Patent Cooperation Treaty or with the National Patent Office
R&D expenditure investment	rei	The total domestic R&D expenditure as a percentage of GDP, including capital and recurrent expenditures of business, government, higher education, and private nonprofit institutions
R&D manpower input	rmi	The number of scientific researchers engaged in research and improvement or development of software for concepts, theories, models, technologies, instruments, and operating methods, expressed as per million
Foreign direct investment	fdi	The proportion of net inflow of foreign direct investment in GDP
Government education expenditure	edu	The proportion of government education expenditure in GDP
GDP per capita	gdp	GDP divided by mid-year population in each country

Based on the previous empirical literature and data availability, three control variables are added to equations (2) and (3), and then three econometric models are considered in this paper, namely, mixed least square models, fixed-effect models, and random-effect models. The specific models are as follows.

2.2.1. Mixed Least Square Models.

$$lnsja_{it} = \alpha_0 + \alpha_1 \times lnrei_{it} + \alpha_2 \times lnrmi_{it} + \alpha_3 \times lnfdi_{it} + \alpha_4 \times lnedu_{it} + \alpha_5 \times lngdp_{it} + \lambda_{it}, \quad (4)$$

$$lnpat_{jt} = \beta_0 + \beta_1 \times lnrei_{jt} + \beta_2 \times lnrmi_{jt} + \beta_3 \times lnfdi_{jt} + \beta_4 \times lnedu_{jt} + \beta_5 \times lngdp_{jt} + \eta_{jt}. \quad (5)$$

2.2.2. Fixed-Effect Models.

$$lnsja_{it} = \gamma_i + \alpha_1 \times lnrei_{it} + \alpha_2 \times lnrmi_{it} + \alpha_3 \times lnfdi_{it} + \alpha_4 \times lnedu_{it} + \alpha_5 \times lngdp_{it} + \lambda_{it}, \quad (6)$$

$$lnpat_{jt} = \nu_j + \beta_1 \times lnrei_{jt} + \beta_2 \times lnrmi_{jt} + \beta_3 \times lnfdi_{jt} + \beta_4 \times lnedu_{jt} + \beta_5 \times lngdp_{jt} + \eta_{jt}. \quad (7)$$

2.2.3. Random-Effect Models.

$$lnsja_{it} = \gamma_i + \alpha_1 \times lnrei_{it} + \alpha_2 \times lnrmi_{it} + \alpha_3 \times lnfdi_{it} + \alpha_4 \times lnedu_{it} + \alpha_5 \times lngdp_{it} + \lambda_{it} + \varepsilon_{it}, \quad (8)$$

$$lnpat_{jt} = \nu_j + \beta_1 \times lnrei_{jt} + \beta_2 \times lnrmi_{jt} + \beta_3 \times lnfdi_{jt} + \beta_4 \times lnedu_{jt} + \beta_5 \times lngdp_{jt} + \eta_{jt} + \mu_{jt}, \quad (9)$$

where lnfdi means logarithm of the foreign direct investment, lnedu denotes logarithm of the government education expenditure, lngdp represents logarithm of the GDP per capita, i and j both mean the countries, t is the period, α and β are both the regression coefficients, $\gamma_i, \nu_j\,(i, j = 1, \ldots, 18)$ represent the unknown intercepts for each country (18 country-specific intercepts), $\lambda_{it}, \eta_{jt}\,(i, j = 1, \ldots, 18; t = 1, \ldots, 6)$ represent the between-entity error, and $\varepsilon_{it}, \mu_{jt}\,(i, j = 1, \ldots, 18; t = 1, \ldots, 6)$ denote the within-entity error.

Finally, using the robust standard deviation of clustering to estimate the standard deviation, we get the heteroscedasticity robust fixed-effect model. The so-called cluster is a cluster of all the observations in each country at different times. The observations in the same cluster allow for correlation; the observed values from different clusters are not correlated.

3. Model Testing and Discussion

3.1. Unit Root Test and Cointegration Test. Considering that sometimes some economic data of time series themselves may not be directly related, often also show a common trend of change and appear the phenomenon of false regression, it is necessary to conduct the stationarity test on the panel data before the regression analysis, so as to avoid the problem of false regression and ensure the stationarity of the data. It is found that all of the eight variables show a time trend by drawing. Therefore, the model with the time trend is used to test the unit root of the eight variables, while the first difference variable has no time trend. The model without time trend is used to test the unit root. Let us use three methods to enhance the reliability of the research results, namely, the LLC test, the ADF test, and the PP test. The test results are shown in Table 2.

From the data in Table 2, it can be seen that in the case of the original value test, except that lnrei and lnrmi tests all passed the 1% significance level test, other variables failed; that is, there is a unit root. The p values corresponding to all the statistics in the first-order difference sequence are all less than 1%, indicating that all variables are stationary after the

Table 2: Unit root test results.

Variables	Original value (time trend)			First differences (no time trend)		
	LLC test	ADF test	PP test	LLC test	ADF test	PP test
lnsja	−24.717***	51.671**	83.433***	−3.046***	67.246***	69.978***
lnpat	−8.818***	50.829*	95.098***	−4.851***	109.487***	111.286***
lnrei	−42.601***	63.174***	98.324***	−8.793***	101.697***	108.817***
lnrmi	−26.270***	71.882***	101.403***	−5.109***	76.861***	77.581***
lnfdi	−14.805***	44.254	82.323***	−12.164***	130.025***	133.415***
lnedu	−3.578***	16.831	23.617	−6.695***	77.284***	83.165***
lngdp	−16.775***	39.923	71.300***	−4.198***	59.500***	61.025***

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

first-order difference sequence; that is, all variables are first-order integrals.

In order to further test whether there is a stable long-term relationship between the variables, we use the Kao test method based on the two-step test of Engle and Granger to test the cointegration. In the first stage of the Kao test, the regression equation is set as that each individual section has a different intercept term and the same trend term, and then all trend coefficients are set to 0. In the second stage, based on the DF test and ADF test, the residual sequence obtained in the first stage is tested for stability. The test results are shown in Table 3. If the p value is less than 1%, it means that the original assumption of rejecting the Kao test "there is no cointegration relationship," that is, there is a stable long-term equilibrium relationship between variables. Therefore, the model established in this paper is correct and regression estimation can be conducted using panel data.

Table 3: Kao cointegration test results.

	t-statistic	p value
ADF (lnsja)	−2.924390	0.0017
ADF (lnpat)	−3.022997	0.0013

ADF (lnsja) and ADF (lnpat), respectively, represent the cointegration test of lnsja and lnpat with explanatory variables and control variables.

3.2. Transnational Regression of R&D Input and Scientific Journal Articles. According to basic models (4), (6), and (8), we make the regression analysis on the panel data of South Asian and Southeast Asian countries. Table 4 lists the regression results of R&D input and the number of scientific journal articles.

Since the growth cycle of the number of scientific journal articles is generally long, R&D input is related to long-term innovation, and the precise time model between R&D input and scientific journal articles is not determined completely according to the annual data. Therefore, let us consider the influence of R&D input on scientific journal articles in a suitable period of time and then set the time period as 3 years. The explanatory variable is the mean number of scientific journal articles in each period, and the other variables are also the mean values of each period. Moreover, the following system has four models, where model I is the mixed least square model, model II is the random-effect model, model III is the fixed-effect model, and model IV is the heteroscedasticity robust fixed-effect model. The weight of each country in the estimation process is the same but different standard error and time error correlations are allowed for each time period.

Compared with model I, model III is more suitable for this study. Model I does not consider the panel and time dimension when processing data. The main reason for extending the data to a panel data structure is to increase the amount of information in the sample. Although the main evidence comes from cross-sectional variations (between countries or regions), the time series dimension (within countries or regions) provides additional information. Within countries, this information is very useful for the explanatory variable in this paper because R&D input changes dramatically over time. In order to be scientific and rigorous, we first test model III with F-test to determine whether to use model I or model III before selecting the model, and then the Hausman test of model II is used to determine whether the following part should select model II or model III. The significance p value of the F test shows that $p < 0.0001 < 0.05$, so it can be seen that model I is rejected and model III is selected for analysis. Moreover, the significance p value of the Hausman test shows that $p = 0.0194 < 0.05$, suggesting that model II is rejected and model III is selected for analysis. In summary, model III is selected after comprehensive consideration. Considering the problem of heteroscedasticity, the heteroscedasticity robust estimation is carried out for model III, and then model IV is obtained. The comparison between model III and model IV shows that the estimated coefficients of the two models are equal but the p value of the coefficient of model IV after the heteroscedasticity robust treatment is generally larger than that of model III. And the number of * sign of lnrei is reduced from 3 to 2; that is, the explanatory variable becomes less explanatory to the explained variable. Moreover, the adjusted R^2 value increases from 0.788 to 0.823, indicating that the overall fitting degree of the equation of model IV is higher. Therefore, model IV makes a more accurate estimate of R&D input and scientific journal articles in this paper.

According to the regression coefficient and p value ($\alpha_1 = 2.139$, $p < 0.05$) corresponding to lnrei, it means that lnsja increases by 2.139 units for each unit of lnrei but its changing degree is small, that is to say, R&D input has a little impact on the number of scientific journal articles. Moreover, the ** sign shows that lnrei is significantly related to

Table 4: The regression results of R&D input and scientific journal articles.

Variables	Model I	Model II	Model III	Model IV
lnrei	0.703*** (6.93)	0.900*** (7.69)	2.139*** (5.01)	2.139** (2.39)
lnrmi	0.491*** (3.71)	0.907*** (5.54)	1.098*** (5.26)	1.098*** (3.03)
lnfdi	−0.005 (−0.16)	−0.015 (−0.69)	0.057** (−2.28)	−0.057*** (−3.21)
lnedu	0.181 (1.59)	0.163 (2.63)	0.169** (2.60)	0.169* (1.69)
lngdp	0.628*** (7.95)	0.561 (6.26)	0.387*** (3.40)	0.387* (1.74)
_cons (intercept)	−12.74*** (−7.45)	−18.74*** (−7.42)	−39.09*** (−5.95)	−39.09** (−2.74)
R^2	0.9418	0.8518	0.8317	0.8317
AdjustedR^2	0.939	0.845	0.788	0.823
F values	357.51***	117.23***	115.35***	84.00***

The values in parentheses are the standard deviation of the coefficient estimation. * $p < 0.1$; ** $p < 0.05$;*** $p < 0.01$.

lnsja within the set confidence interval. Through lnrmi corresponding regression coefficients and the p value ($\alpha_2 = 1.098$, $p < 0.01$), we can know that the coefficient is positive, which means that lnsja increases by 1.098 units for every 1 unit of lnrmi in the case of other conditions unchanged. But its changing degree is smaller than that of lnrei; that is, the impact of R&D manpower input on scientific journal articles is less than that of R&D expenditure investment. Moreover, the *** sign means that lnrmi within the set confidence interval is very significantly correlated with lnsja. By adding two explanatory variables into the model, it can be concluded that the effect of R&D input on scientific journal articles cannot be considered from a single variable and that both R&D expenditure and manpower input can increase the number of scientific journal articles, so as to contribute to national technology innovation.

In the control variables, the coefficient margins of lnedu and lngdp are significantly positive, which can positively promote the publication of scientific journal articles. However, the coefficient of lnfdi is extremely negative, which will inhibit the production of scientific journal articles. With the increase of foreign investment in most south and southeast Asian countries except Singapore, a large amount of funds are used to purchase the readily available and relatively backward foreign technological achievements. As a result, local innovation subjects in many countries lack energy and funds to engage in innovation activities, which weakens the independent innovation ability of local innovation subjects and reduces the correlation between R&D expenditure and technological innovation. As mentioned above, this is consistent with the negative correlation between foreign investment and the realization of technological innovation when foreign investment is taken as the control variable in this paper.

Through the fixed-effect intercept of each country in the model IV, we can know the specific equation of the relationship between R&D input and scientific journal articles in each country from Table 5, such as the equation of Singapore, Philippines, and Myanmar:

$$\begin{aligned} lnsja_{Singapore} &= 3.37946 + 2.139 \times lnrei_t \\ &\quad + 1.098 \times lnrmi_t - 0.057 \times lnfdi_t \\ &\quad + 0.169 \times lnedu_t + 0.387 \times \mathrm{lngdp}_t + \lambda_t, \end{aligned} \tag{10}$$

$$\begin{aligned} lnsja_{Philippines} &= 0.13289 + 2.139 \times lnrei_t \\ &\quad + 1.098 \times lnrmi_t - 0.057 \times lnfdi_t \\ &\quad + 0.169 \times lnedu_t + 0.387 \times lngdp_t + \lambda_t, \end{aligned} \tag{11}$$

$$\begin{aligned} lnsja_{Myanmar} &= -2.07867 + 2.139 \times lnrei_t \\ &\quad + 1.098 \times lnrmi_t - 0.057 \times lnfdi_t \\ &\quad + 0.169 \times lnedu_t + 0.387 \times lngdp_t + \lambda_t, \end{aligned} \tag{12}$$

where $\lambda_t = -39.09$.

Table 5 shows the fixed-effect intercepts of each country in model IV, indicating that South and Southeast Asian countries have significant differences in scientific journal articles. From a horizontal perspective, India, Malaysia, Singapore, and Thailand have the largest number of scientific journal articles in South Asian and Southeast Asian countries, followed by Pakistan, Indonesia, Vietnam, Sri Lanka, Philippines, Bangladesh, and Brunei. The number of scientific journal articles in Cambodia, Lao PDR, Myanmar, Nepal, Bhutan, Timor-Leste, and Maldives is lagging behind, which further shows that these countries need to increase R&D expenditure investment and manpower input reasonably to improve the level of technological innovation.

3.3. Transnational Regression of R&D Input and Patent Applications. According to the basic model (5), (7), and (9), we make the regression analysis on the panel data of South Asian and Southeast Asian countries. Table 6 lists the regression results of R&D input and patent application.

As the growth cycle of patent applications quantity is generally long, R&D input is related to long-term technology innovation, and the precise time model between R&D input and patent applications is not determined completely according to the annual data. Therefore, let us consider the effect of R&D input on patent applications in a moderately long period of time and then set the time period as 3 years. The explanatory variable is the mean number of patent applications in each period and the other variables are also the mean values of each period. Moreover, the following system has five models, where model V is the mixed least square model, model VI is the random-effect model, model

Table 5: The fixed-effect intercepts of each country in model IV.

Country	Intercept	Country	Intercept	Country	Intercept
India	4.54196	Vietnam	0.90077	Lao PDR	−1.94523
Malaysia	4.33915	Sri Lanka	0.43807	Myanmar	−2.07867
Singapore	3.37946	Philippines	0.13289	Nepal	−2.29419
Thailand	3.20740	Bangladesh	−0.35921	Bhutan	−2.68533
Pakistan	1.87514	Brunei	−0.54005	Timor-Leste	−4.39069
Indonesia	1.84323	Cambodia	−1.93539	Maldives	−4.42933

Table 6: The regression results of R&D input and patent applications.

Variables	Model V	Model VI	Model VII	Model VIII	Model IX
lnrei	1.022*** (7.09)	1.015*** (8.00)	1.277*** (3.1)	1.277*** (3.55)	1.241*** (4.11)
lnrmi	1.315*** (6.85)	1.062*** (6.27)	0.842*** (4.18)	0.842*** (2.79)	0.860*** (3.70)
lnfdi	−0.061 (−1.59)	−0.037* (−1.72)	−0.031 (−1.29)	−0.031 (−1.26)	—
lnedu	0.318** (2.20)	0.117* (1.91)	0.113* (1.80)	0.113* (1.58)	0.135* (2.06)
lngdp	0.292** (2.31)	0.144 (1.58)	0.070 (0.64)	0.070 (0.44)	—
_cons(intercept)	−25.17*** (−8.86)	−21.81*** (−8.17)	−23.90*** (−3.77)	−23.90*** (−3.66)	−23.34*** (−4.85)
R^2	0.9253	0.7368	0.6148	0.6148	0.6031
Adjusted R^2	0.922	0.724	0.515	0.596	0.592
F values	119.08***	57.12***	27.13***	16.25***	15.26***

The values in parentheses are the standard deviation of the coefficient estimation. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

VII is the fixed-effect model, and both model VIII and model IX are the heteroscedasticity robust fixed-effect models. The weight of each country in the estimation process is the same but different standard error and time error correlations are allowed for each time period.

As mentioned above, for the sake of scientificity and rigor, the F test of model VII is conducted first before selecting the model, and then the Hausman test of model VI is conducted to determine whether the following model should be model V, model VI, or model VII. Therefore, the significance p value of the F test shows that $p < 0.0001 < 0.05$, so it can be seen that model VII is selected for analysis. Moreover, the significance p value of the Hausman test shows that $p = 0.0207 < 0.05$, suggesting that model VII is also selected for analysis. In a word, model VII is selected after comprehensive consideration. Compared with model V and model VI in this paper, model VII is the most suitable for this study. Considering the problem of heteroscedasticity, the heteroscedasticity robust estimation is carried out for model VII and model VIII is obtained. The comparison between model VII and model VIII shows that the coefficient values estimated by the two models are equal but the p value of the coefficient of model VIII after the heteroscedasticity robust treatment is larger than that of model VII. Moreover, the adjusted R^2 value of the two models increases from 0.515 to 0.596, indicating that the overall fitting degree of the equation of model VIII is higher. Therefore, model VIII makes a more accurate estimate of R&D input and patent applications in this paper.

Because lnfdi and lngdp of the control variables in model VIII have no significant effect on patent applications ($p > 0.1$), we get model IX after eliminating these two variables. And model IX is still the heteroscedasticity robust fixed-effect model. According to the regression coefficient and p value ($\beta_1 = 1.241$, $p < 0.01$) corresponding to lnrei, it means that lnpat increases by 1.241 units for each unit of lnrei but its changing degree is small; that is to say, R&D input has little impact on patent applications quantity. Moreover, the *** sign shows that lnrei is very significantly related to lnpat within the set confidence interval. Through lnrmi corresponding regression coefficients and the p value ($\beta_2 = 0.86$, $p < 0.01$), we can know that the coefficient is positive, which means that lnpat increases by 0.86 units for each unit of lnrmi in the case of other conditions unchanged. But its changing degree is smaller than that of lnrei; that is, the impact of R&D manpower input on patent application quantity is less than that of R&D expenditure investment. Moreover, the *** sign means that lnrmi within the set confidence interval is also very significantly correlated with lnpat. By adding two explanatory variables into the model, it can be concluded that the impact of R&D input on patent applications cannot be considered from a single variable and both R&D expenditure and manpower input can increase patent application quantity, so as to promote national innovation.

Through the fixed-effect intercept of each country in model IX, we can know the specific equation of the relationship between R&D input and patent applications in each country from Table 7, such as the equation of Singapore, Philippines, and Myanmar:

$$lnpat_{Singapore} = 1.65928 + 1.241 \times lnrei_t + 0.86 \times lnrmi_t + 0.135 \times lnedu_t - \eta_t, \quad (13)$$

$$lnpat_{Philippines} = 0.7860 + 1.241 \times lnrei_t + 0.86 \times lnrmi_t + 0.135 \times lnedu_t - \eta_t, \quad (14)$$

$$lnpat_{Myanmar} = -3.1008 + 1.241 \times lnrei_t + 0.86 \times lnrmi_t + 0.135 \times lnedu_t - \eta_t, \quad (15)$$

where $\eta_t = -23.34$.

Table 7: The fixed-effect intercepts of each country in model IX.

Country	Intercept	Country	Intercept	Country	Intercept
Singapore	1.65928	Philippines	0.78600	Bhutan	−0.75375
India	1.50356	Brunei	0.39971	Maldives	−0.77819
Malaysia	1.32618	Sri Lanka	0.16256	Bangladesh	−0.85570
Indonesia	1.20668	Pakistan	−0.05891	Cambodia	−1.15125
Thailand	1.14805	Nepal	−0.20206	Lao PDR	−1.94089
Vietnam	1.06440	Timor-Leste	−0.41490	Myanmar	−3.10078

Table 7 shows the fixed-effect intercepts of each country in model IX, indicating that South and Southeast Asian countries have significant differences in patent applications. From a horizontal perspective, Singapore, India, Malaysia, Indonesia, and Thailand have the largest number of scientific journal articles in South Asian and Southeast Asian countries, followed by Vietnam, Philippines, Brunei, Sri Lanka, Pakistan, and Nepal. The number of scientific journal articles in Timor-Leste, Bhutan, Maldives, Bangladesh, Cambodia, Lao PDR, and Myanmar is lagging behind, which further shows that these countries need to increase R&D expenditure investment and manpower input reasonably to improve the level of technological innovation.

4. Model Result Analysis

4.1. R&D Input in South Asia and Southeast Asia Significantly Promotes Technological Innovation. Based on the regression analysis above, we find that both R&D expenditure and manpower input in 18 countries are significantly positively correlated with the number of scientific journal articles and patent applications, so R&D input significantly promotes technology innovation. The reasons may be as follows. Firstly, with the rapid development of science in South and Southeast Asia, the complexity and uncertainty of technology innovation increase, and the difficulty of advancing the development of frontier science increases. In this context, R&D input in various countries can reduce the R&D cost and risk and promote the motivation of innovation for some R&D institutions, especially in Cambodia, Lao PDR, Myanmar, Timor-Leste, Bangladesh, Bhutan, Maldives, and Nepal. Secondly, R&D expenditure investment provides more financial support for the innovation output of South Asian and Southeast Asian countries. And it can attract more funds into the field of scientific research and expand the scale of scientific research expenditures, either directly by promoting enterprises to increase their own investment in technology, or indirectly by guiding financial institutions to make loans, especially in Singapore, India, Malaysia, Indonesia, and Thailand. These methods promote technology introduction and equipment upgrading, providing more favorable material conditions for R&D activities. Thirdly, from the perspective of the long-term development needs of enterprises and related institutions, they need to increase the R&D manpower input. And R&D manpower input also provides sufficient manpower support for the innovation output of South Asian and Southeast Asian countries to ensure the smooth development of R&D activities and encourages researchers to improve their own strength in the hope of publishing more high-quality scientific articles and patents.

4.2. Efficiency of R&D Input Promoting Technology Innovation Is Low and Needs to Be Improved. The regression analysis above indicates that the positive effect of R&D expenditure and manpower input on the number of scientific journal articles and patent applications is both a little weak within the established confidence interval, among which the positive impact of R&D manpower input is generally weaker than that of expenditure investment. So, we think that the effect of R&D expenditure and manpower input on promoting innovation is not obvious and the using efficiency of R&D input is low. In particular, the using efficiency of the R&D manpower input in South Asia and Southeast Asia has more needs to be improved.

The reasons may be as follows. (a) The unreasonable R&D expenditure structure: In the high-speed stage of economic development, many countries in South Asia and Southeast Asia prefer to invest in applied research and enterprise product development with strong competitiveness and higher returns, so that they cannot guarantee sufficient investment in basic research. It will lead to the brain drain, the lack of enthusiasm for innovation, and the difficulty in national innovation. (b) The inappropriate R&D expenditure investment scale: At present, most of the national innovation systems of many countries in South Asia and Southeast Asia are still in the initial stage and the achievements of technological innovation are highly sensitive to the scale of the R&D expenditure investment. So, the scale of the R&D expenditure investment is not large enough, which will limit the process of innovation to some extent. This is especially true in countries where technological innovation is more backward, such as in Cambodia, Lao PDR, Myanmar, Timor-Leste, Bangladesh, Bhutan, Maldives, Nepal, and so on. (c) The long-term and post-economic problems of the R&D manpower input: On the one side, the R&D manpower input needs the transformation of knowledge and the time to carry out the activities of technological innovation, which makes the enterprises and institutions that pay attention to the current benefits less willing to invest human capital. On the other side, due to the ownership of human capital and the mobility of human resources, enterprises and social institutions are not willing to invest a lot of manpower. (d) The "quality" and "quantity" problems of the R&D manpower input: Nowadays, the evaluation index of scientific researchers in South Asia and

Southeast Asia is more focused on academic achievements, which brings about the lack of motivation for scientific researchers to industrialize and market their achievements. Universities and research institutes have a weak connection with the industry and their research achievements are out of step with the market demand. Although there are more and more technological personnel in South Asia and Southeast Asia, there is still a shortage of high-level innovative talents who can truly make technological breakthroughs.

4.3. South Asian and Southeast Asian Countries Are Divided into Three Levels according to the Fixed-Effect Intercepts of Each Country. Comprehensively considering the fixed-effect intercepts of each country in the heteroscedasticity robust fixed-effect models of scientific journal articles and patent applications, South Asian and Southeast Asian countries can be roughly divided into three levels in terms of technological innovation. Countries with higher technological innovation levels are Singapore, India, Malaysia, and Thailand; the countries with a medium level of technological innovation are Indonesia, Vietnam, Philippines, Brunei, Sri Lanka, and Pakistan; and the countries with a low level of technological innovation are Cambodia, Lao PDR, Myanmar, Timor-Leste, Bangladesh, Bhutan, Maldives, and Nepal. Through the equations (10) and (13), the equation of the relationship between R&D investment and technological innovation in Singapore is obtained, which provides some inspiration and helps in the optimization of R&D input in developed countries. Through the equations (11) and (14), the equation of the relationship between R&D input and technological innovation in Philippines is obtained, which provides a reference for developing countries to formulate technology policies. Through the equations (12) and (15), the equation of the relationship between R&D input and technological innovation in Myanmar is obtained, which can find some reasons for the relatively backward countries in technological innovation and bring some new inspiration to these countries.

5. Policy Proposals

5.1. All Countries in South Asia and Southeast Asia Should Continue to Increase the R&D Expenditure and Manpower Input to Promote Technology Innovation. On the macrolevel, on the one side, all countries in South Asia and Southeast Asia should continue to increase manpower input in innovation and scientific research activities. On the other side, it is necessary to increase the input intensity in the R&D expenditure and actively guide enterprises and the society to increase it, which can enhance the national R&D input level, ensure the continuous growth of national innovation output, and provide strong support for the national technological innovation. On the microlevel, enterprises need to increase the input in high-quality personnel for improving the innovation performance, and the most important thing is to increase the R&D expenditure investment. The R&D institutions should not increase manpower input on a large scale but should keep the R&D expenditure investment moderately stable, stimulate the innovation enthusiasm of current direct manpower, and optimize the structure of manpower input. Colleges and universities should maintain a stable R&D input, on this basis, appropriately increasing the intensity of manpower input. Therefore, increasing the R&D expenditure and manpower input is the basic means to improve the level of innovation.

5.2. Countries Should Improve the Efficiency of the R&D Input to Promote Technology Innovation. Improving the efficiency of the R&D input is the basic means to increase R&D input to promote the effect of technology innovation. Based on the above analysis, in terms of the structure of expenditure investment, all countries need to make more scientific and reasonable decision-making basis in terms of how to rationally allocate expenditures in basic research, applied research, and enterprise product development. In terms of expenditure investment scale, as a key factor that directly affects the output of innovation, the R&D expenditure investment should continue to increase its investment scale, expand the breadth and depth of technology and knowledge spillover, reasonably optimize the staged allocation of the R&D expenditure, and especially increase the scale of investment in the transformation of technology achievements.

Aiming at the problem of R&D manpower input, the R&D manpower input in South Asia and Southeast Asia not only focuses on the increase in the number of the R&D personnel to maintain a proper proportion relationship with the R&D expenditure, but also pays attention to the improvement of R&D personnel's knowledge reserve and professional skill. In particular, it is necessary to give full play to the technical expertise of core technical personnel in the transformation of technical achievements and the marketization of products, so as to promote the scale of national innovation output increasing returns. In particular, the technologically backward countries of South Asia and Southeast Asia should do these better, especially in Cambodia, Lao PDR, Myanmar, Timor-Leste, Bangladesh, Bhutan, Maldives, and Nepal.

In the high-speed stage of economic development, the R&D manpower in various countries is concentrated in the aspect of experimental development, which is more suitable for the current development situation. However, basic research is the cornerstone of the long-term and stable development of science and plays a supporting role in the overall development of R&D activities. Gradually strengthening basic research is very significant for the long-term improvement of national innovation capacity and the development of technology, especially in Singapore, India, Malaysia, Indonesia, and Thailand. So, it is hoped that relevant departments of all countries in South Asia and Southeast Asia will take appropriate measures to provide enterprises and society with a more leading understanding of basic research and increase input in basic research.

Data Availability

The basic data of this article can be downloaded from the official website of the World Development Indicators of the

World Bank. There is a public dataset that may be downloaded from https://databank.worldbank.org/home. Data supporting the results of this study can be obtained from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was sponsored by the National Natural Science Foundation of China (NSFC) (71163046) and the Natural Science Foundation of Yunnan Province (2018RD004).

References

[1] Z. Griliches, "Issues in assessing the contribution of research and development to productivity growth," *The Bell Journal of Economics*, vol. 10, no. 1, pp. 92–116, 1979.

[2] A. B. Jaffe, "Real affects of academic research," *American Economics Review*, vol. 79, no. 5, pp. 957–970, 1989.

[3] P. M. Romer, "Endogenous technological change," *Journal of Political Economy*, vol. 98, no. 5, pp. 71–102, 1990.

[4] C. I. Jones, "R&D-based models of economic growth," *Journal of Political Economy*, vol. 103, no. 4, pp. 759–784, 1995.

[5] F. M. Scherer, "Firm size, market structure, opportunity and the output of patented inventious," *American Economic Review*, vol. 55, no. 5, pp. 1097–1126, 1956.

[6] A. Pake and Z. Griliches, "Patents and R&D at the firm level: a first look report," *Economics Letters*, vol. 5, no. 4, pp. 377–381, 1980.

[7] B. Crepon, E. Duguet, J. Mairessec et al., "Research, innovation and productivity: an econometric analysis at the firm level," *Economics of Innovation and New Technology*, vol. 7, no. 2, pp. 115–158, 1998.

[8] W. Nasierowski and F. J. Arcelus, "On the efficiency of national innovation systems," *Socio-Economic Planning Sciences*, vol. 3, no. 37, pp. 215–234, 2003.

[9] G. Pellegrino, M. Piva, and M. Vivarelli, "Young firms and innovation: a microeconometric analysis," *Structural Change and Economic Dynamics*, vol. 23, no. 4, pp. 329–340, 2012.

[10] F. Bogliacino and M. Pianta, "Profits, R&D, and innovation-a model and a test," *Industrial and Corporate Change*, vol. 22, no. 3, pp. 649–678, 2013.

[11] J. Baumann and A. S. Kritikos, "The link between R&D, innovation and productivity: are micro firms different?" *Research Policy*, vol. 45, no. 6, pp. 1263–1274, 2016.

[12] H. L. Zhu, S. L. Zhao, and A. Abbas, "Relationship between R&D grants, R&D investment, and innovation performance: The moderating effect of absorptive capacity," *Journal of Public Affairs*, vol. 20, no. 1, p. 1973, 2020.

[13] G. H. Jefferson, B. Huamao, X. J. Guan et al., "R&D performance in Chinese industry," *Economics of Innovation & New Technology*, vol. 15, no. 4-5, pp. 345–366, 2002.

[14] S. A. Le, B. Walters, and M. Kroll, "The moderating effects of external monitors on the relationship between R&D spending and firm performance," *Journal of Business Research*, vol. 59, no. 2, pp. 278–287, 2006.

[15] D. Aboody and B. Lev, "Information asymmetry, R&D, and insider gains," *The Journal of Finance*, vol. 55, no. 6, pp. 2747–2766, 2000.

[16] S. David Young, *EVA and Value Based Management: A Practical Guide to Implementation*, Mcgraw Hill Professional, New York, NY, USA, 2000.

[17] P. A. Abetti, "Critical success factors for radical technological innovation: a five case study," *Creativity and Innovation Management*, vol. 9, no. 4, pp. 208–221, 2000.

[18] M. A. Hit, R. D. Ireland, J. S. Harrison et al., "Effects of acquisitions on R&D inputs and outputs," *Academy of Management Journal*, vol. 34, no. 3, pp. 693–706, 1991.

[19] C. Shi, "On the trade-off between the future benefits and riskiness of R&D: a bondholders' perspective," *Journal of Accounting and Economics*, vol. 35, no. 2, pp. 227–254, 2003.

[20] G. Bin, "Firm size, R&D, and performance: an empirical analysis on software industry in China," *Science Research Management*, vol. 1, pp. 613–616, 2006.

[21] P. Koellinger, "The relationship between technology, innovation, and firm performance-empirical evidence from e-business in Europe," *Research Policy*, vol. 37, no. 8, pp. 1317–1328, 2008.

[22] Y. M. Lu and C. M. Wang, "Effect of R&D investment on performance of Chinese listed companies—take manufacturing and IT industry as an example," *Science and Technology Management Research*, vol. 31, no. 5, pp. 122–127, 2011.

[23] X. Y. Qu, P. Lu, H. X. Wang et al., "Empirical analysis of the impact of R&D investment on the transformation and upgrading of China's traditional manufacturing industry," *Statistics and Decision-Making*, vol. 54, no. 5, pp. 120–123, 2020.

[24] C.-P. Chen, J.-L. Hu, C.-H. Yang et al., "An international comparison of R&D efficiency of multiple innovative outputs: the role of the national innovation system," *Innovation*, vol. 13, no. 3, pp. 341–360, 2011.

[25] E. C. Wang and W. Huang, "Relative efficiency of R&D activities: a cross-country study accounting for environmental factors in the DEA approach," *Research Policy*, vol. 36, no. 2, pp. 260–273, 2007.

[26] A. Ashish, F. Andrea, and G. Alfonso, "Markets for technology and their implications for corporate strategy," *Industrial and Corporate Change*, vol. 10, no. 2, pp. 419–451, 2001.

[27] M. Fritsch and G. Franke, "Innovation, regional knowledge spillovers and R&D cooperation," *Research Policy*, vol. 33, no. 2, pp. 245–255, 2004.

[28] M. Buesa, J. Heijs, and T. Baumert, "The determinants of regional innovation in Europe: a combined factorial and regression knowledge production function approach," *Research Policy*, vol. 39, no. 6, pp. 722–735, 2010.

[29] Z. J. Acs, L. Anselin, and A. Varga, "Patents and innovation counts as measures of regional production of new knowledge," *Research Policy*, vol. 31, no. 7, pp. 1069–1085, 2002.

[30] W. Zhong, W. Yuan, S. X. Li, and Z. Huang, "The performance evaluation of regional R&D investments in China: an application of DEA based on the first official China economic census data," *Omega*, vol. 39, no. 4, pp. 447–455, 2011.

[31] W. M. Cohen, R. R. Nelson, and J. P. Walsh, "Links and impacts: the influence of public research on industrial R&D," *Management Science*, vol. 48, no. 1, pp. 1–23, 2002.

[32] M. G. Kocher, M. Luptacik, and M. Sutter, "Measuring productivity of research in economics: a cross-country study using DEA," *Socio-Economic Planning Sciences*, vol. 40, no. 4, pp. 314–332, 2006.

[33] A. Hashimoto and S. Haneda, "Measuring the change in R&D efficiency of the Japanese pharmaceutical industry," *Research Policy*, vol. 37, no. 10, pp. 1829–1836, 2008.

[34] V. Thomas, S. Sharma, and S. K. Jain, "Using patents and publications to assess R&D efficiency in the states of the USA," *World Patent Information*, vol. 33, no. 1, pp. 4–10, 2011.
[35] U. Han, M. Asmild, and M. Kunc, "Regional R&D efficiency in Korea from static and dynamic perspectives," *Regional Studies*, vol. 50, no. 7, pp. 1170–1184, 2016.
[36] S.-H. Yoo, "Public R&D expenditure and private R&D expenditure: a causality analysis," *Applied Economics Letters*, vol. 11, no. 11, pp. 711–714, 2004.
[37] H. Y. Lee and Y. T. Park, "An international comparison of R&D efficiency: DEA approach," *Asian Journal of Technology Innovation*, vol. 13, no. 2, pp. 207–222, 2005.
[38] S. S. J. Thomas, "Inter-country R&D efficiency analysis: an application of data envelopment analysis," *Scientometrics*, vol. 76, no. 3, p. 483, 2008.
[39] M. Docekalova and N. Bockova, "The use of data envelopment analysis to assess the R&D Effectiveness of the Czech manufacturing industry," *Verslas: Teorija Ir Praktika*, vol. 14, no. 4, pp. 308–314, 2013.
[40] K. H. Chen and M. T. Kou, "Staged efficiency and its determinants of regional innovation systems: a two-step analytical procedure," *Annals of Regional Science*, vol. 52, no. 2, pp. 627–657, 2014.
[41] O. Ryoji, "Innovation, foreign direct investment and local content requirement," *Osaka City University Economic Review*, vol. 45, pp. 35–52, 2009.
[42] V. N. Balasubramanyam, D. Sapsford, and M. A. Salisu, "Foreign direct investment and growth," *Social Science Electronic Publishing*, vol. 70, no. 3, pp. 291–314, 1998.
[43] J. M. Arnold and B. S. Javorcik, "Gifted kids or pushy parents? foreign direct investment and plant productivity in Indonesia," *Journal of International Economics*, vol. 79, no. 1, pp. 42–53, 2010.
[44] K. Chen and J. Guan, "Measuring the efficiency of China's regional innovation systems: application of network data envelopment analysis (DEA)," *Regional Studies*, vol. 46, no. 3, pp. 355–377, 2012.
[45] A. M. P. Feldman, "R&D spillovers and the geography of innovation and production," *American Economic Review*, vol. 86, no. 3, pp. 630–640, 1996.
[46] J. H. Grossman, P. P. Reid, and R. P. Morgan, "Contributions of academic research to industrial performance in five industry Sectors," *Journal of Technology Transfer*, vol. 26, no. 1-2, pp. 143–152, 2001.
[47] K. Kim, "An analysis of source of growth in East Asian economies and R&D spillover effects," *Korea & the World Economy*, vol. 1, no. 1, pp. 83–107, 2000.
[48] M. Jerzmanowsk, "Total factor productivity differences: appropriate technology vs. efficiency," *European Economic Review*, vol. 51, no. 8, pp. 2080–2110, 2007.
[49] T. Wei, "Impact of energy efficiency gains on output and energy use with Cobb-Douglas production function," *Energy Policy*, vol. 35, no. 4, pp. 2023–2030, 2007.

Hindawi
Discrete Dynamics in Nature and Society
Volume 2020, Article ID 4568171, 9 pages
https://doi.org/10.1155/2020/4568171

Research Article

A Novel Intelligent Recommendation Algorithm Based on Mass Diffusion

Guanglai Tian,[1] **Shuang Zhou**,[1] **Gengxin Sun**,[1] **and Chih-Cheng Chen**[2,3]

[1]*School of Data Science and Software Engineering, Qingdao University, Qingdao 266071, China*
[2]*Department of Industrial Engineering and Management, Chaoyang University of Technology, Taichung 413310, Taiwan*
[3]*Information and Engineering College, Jimei University, Xiamen, Fujian 361021, China*

Correspondence should be addressed to Gengxin Sun; sungengxin@qdu.edu.cn and Chih-Cheng Chen; 3343033397@qq.com

Received 1 June 2020; Revised 17 October 2020; Accepted 31 October 2020; Published 16 November 2020

Academic Editor: Chi-Hua Chen

Social recommendation algorithm is a common tool for recommending interesting or potentially useful items to users amidst the sea of online information. The users usually have various relationships, each of which has its unique impact on the recommendation results. It is unlikely to make accurate recommendations solely based on one relationship. Based on user-item bipartite graph, this paper establishes a multisubnet composited complex network (MSCCN) of multiple user relationships and then extends the mass diffusion (MD) algorithm into a novel intelligent recommendation algorithm. Two public online datasets, namely, Epinions and FilmTrust, were selected to verify the effect of the proposed algorithm. The results show that the proposed intelligent recommendation algorithm with two types of relationships made much more accurate recommendations than that with a single relationship and the traditional MD algorithm.

1. Introduction

The era of information technology is defined by the explosion of all sorts of information. The wealth of information not only provides human with more choices but also causes serious information overload [1–5]; amidst the sea of information, it is difficult for users to find the information that they really need. As a result, how to provide users with useful information from massive data has become a research hotspot. To solve information overload, many recommendation systems have been developed to predict user preference based on their historical data and recommend interesting or potentially useful information to them [6–8].

The common recommendation methods can be divided into collaborative filtering (CF) recommendation [9], content-based recommendation [10], spectrum analysis recommendation [11, 12], and hybrid recommendation [13, 14]. The recommendation methods based on user or item similarity are also quite popular, namely, mass diffusion (MD) algorithm and heat spreading (HeatS) algorithm. The MD is a classic algorithm capable of making accurate recommendations of items to users. By the principle of energy distribution, the algorithm spreads energy from the initial items to different items and recommend the items with relatively high energy at a high probability.

After predicting user preference, Bin et al. [15] evaluated the popularity of each item and used it to modify the MD algorithm. Based on mass diffusion and heat spreading in physical kinetics, Hu et al. [16] summarized the mechanism of energy distribution and realized a recommendation mechanism based on the tripartite of user, item, and topic. Li et al. [17] managed to improve recommendation quality through singular value decomposition (SVD) of the recommendation results.

The above studies mainly recommend items to users based on the item scores rated by users. However, none of them have considered the influence of multiple social relationships on user preference. In fact, the preference of a user is affected by his/her trusted friends. Therefore, this paper constructs a complex network with multiple social relationships. The network incorporates multiple user

relationships into the recommendation system, which effectively improves the recommendation accuracy.

2. Multisubnet Composited Complex Network (MSCCN) Model

There are various systems that can be modeled in our daily life, such as the urban system, ecosystem, transport system, and recommendation system. In a complex system, each object can be treated as a node, and each relationship between objects can be represented as an edge, turning the system into a complex network [18–20].

The MSCCN model [20–23] can describe multiple relationships among heterogeneous individuals in complex systems. For example, users and items can be considered nodes, while the user relationships and user-item relationships can be regarded as edges in the recommendation system.

By integrating multiple subnets and relationships among complex systems, the MSCCN model can be expressed as a four-tuple $G = (V, E, R, F)$, where $V = \{v_1, v_2, \ldots, v_m\}$ is the set of $m = |V|$ nodes, $E = \{v_h, v_l | v_h, v_l \in V, 1 \le h, l \le m\} \subseteq V \times V$ is the set of edges among the nodes, $R = R_1 \times \cdots \times R_i \times \cdots \times R_n = \{(r_1, \ldots, r_i, \ldots, r_n) \mid r_i \in R_i, 1 \le i \le n\}$ is the set of relationships (r_i is a type of relationship; n is the number of relationship types), and F: $E \longrightarrow^{\varphi} R$ is the mapping of set E to R by the φ function, reflecting the type of relationship on each edge.

The structure of a typical MSCNN model is presented in Figure 1, where $R = R_1 \times R_2$, $R_1 = \{r_1\}$, and $R_2 = \{r_2\}$. As shown in Figure 1, the edges v_1, v_2, v_1, v_3, and v_2, v_3 only have R_1 relationship, the edges v_2, v_5, v_3, v_5, and v_4, v_5 only have R_2 relationship, and the edge v_3, v_4 has both R_1 and R_2 relationships.

3. MD Algorithm Based on MSCCN Model

3.1. MD Algorithm. As proposed by Li et al. [17], the MD algorithm can be implemented as follows: First, a unit of energy is placed on the items selected by the target user. Then, the energy spreads from the items to the users who select them. At this time, the energy h_i obtained by user u_i can be calculated by

$$h_i = \sum_{\beta=1}^{n} \frac{a_{i\beta} f_\beta}{k_\beta}, \tag{1}$$

where k_β is the degree of item o_β; n is the number of items; f_β is the energy of item o_β; $a_{i\beta}$ is the selection variable of items. If item o_β is selected by user u_i, then $a_{i\beta} = 1$.

Next, the energy that is spread to users is distributed to the items selected by users. After the distribution, the energy f'_α obtained by item o_α can be expressed as

$$f'_\alpha = \sum_{i=1}^{m} \frac{a_{i\alpha} h_i}{d_i}, \tag{2}$$

where d_i is the degree of user u_i; m is the number of users.

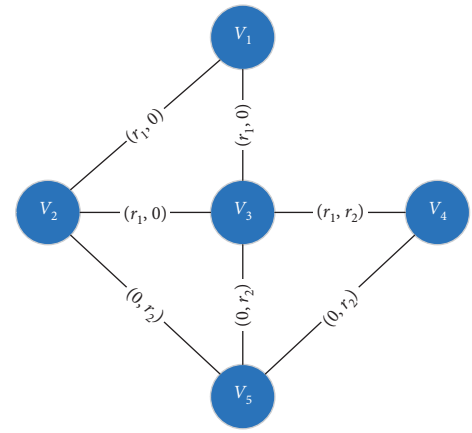


Figure 1: The structure of a typical MSCCN model.

Let $\vec{f}$ and $\vec{f}'$ be the initial and final energy of items, respectively. Then, the entire diffusion process can be defined as

$$\vec{f}' = W^M \vec{f}, \tag{3}$$

where W^M is the state transition matrix. Each element $w^M_{\alpha\beta}$ of the matrix satisfies

$$w^M_{\alpha\beta} = \frac{1}{k_\beta} \sum_{i=1}^{m} \frac{a_{i\alpha} a_{i\beta}}{d_i}. \tag{4}$$

After diffusion, the unselected items are arranged in descending order by the resource quantity, forming a recommendation list. This recommendation list is generated solely based on user-item relationships, without considering the social relationships of the target user.

The workflow of the MD algorithm is illustrated in Figure 2. It can be inferred that $\vec{f} = (1, 0, 1, 0)$ is the initial energy of items and $\vec{f}' = (19/24, 5/24, 5/8, 3/8)$ is the final energy of items after being processed by MD algorithm.

3.2. MD Algorithm Based on MSCCN Network (SMD). By the MD algorithm, the recommendation list is generated solely based on user-item relationships, failing to consider the social relationships of the target user. To overcome the defect, this paper introduces the social relationship into the recommendation system by setting up a multirelationship network, allowing the initial energy of item nodes to spread along the edges in the network.

Let r_1 and r_2 be user-item relationship and user relationship in the multirelationship network, respectively. It is assumed that the weights of the two types of relationship are both 1. Then, the proportionality coefficient sf_1 of r_1 and that sf_2 of r_2 satisfy $sf_1 + sf_2 = 1$. Assuming that sf_1 is p, then $sf_2 = 1 - p$, $p \epsilon (0, 1)$. After introducing one type of social relationship, the SMD algorithm can be implemented as follows:

(i) *Step 1.* One unit of energy is distributed to the items selected by the target user, serving as the initial energy of items.

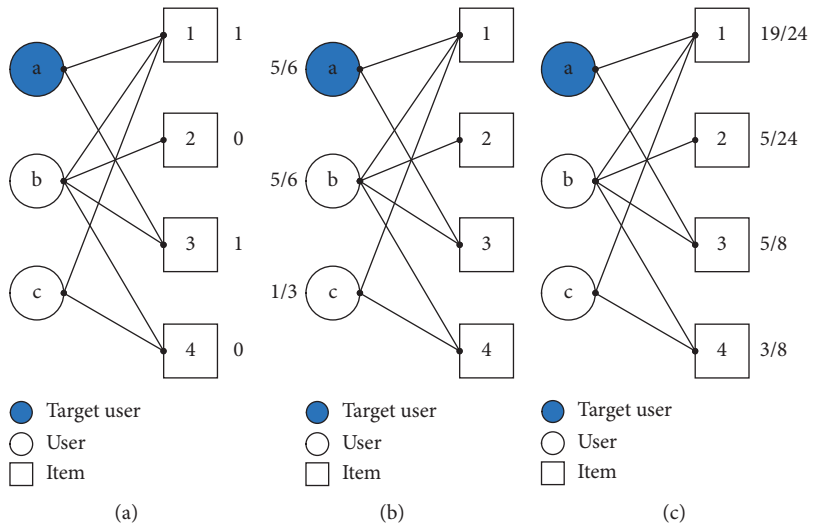


Figure 2: The MD process based on bipartite graph.

(ii) *Step 2*. The initial energy of items spread to the users who have selected them. The energy h_i obtained by user u_i can be described as

$$h_i = \sum_{\beta=1}^{n} \frac{a_{i\beta}^1 f_\beta}{\widehat{k}_{v_\beta}^{r_1}}. \tag{5}$$

where $a_{i\beta}^1$ and $a_{i\beta}^2$ are selection variables of items; $\widehat{k}_{v_\beta}^{r_1}$ is the degree of item o_β for relationship r_1. If $a_{i\beta}^1 = 1$, user u_i and item o_β have r_1 relationship. If $a_{i\beta}^2 = 1$, user u_i and item o_β have r_2 relationship.

(iii) *Step 3*. The energy spreads to users and items along edges in the composite network. Then, the energy h_j' obtained by user u_j and that f_α' obtained by item o_α can be, respectively, described as

$$\begin{aligned} h_j' &= \sum_{i=1}^{m} \left(\frac{a_{ij}^1 sf_1 h_i}{\widehat{k}_{v_i}^{r_1}} + \frac{a_{ij}^2 sf_2 h_i}{\widehat{k}_{v_i}^{r_2}} \right), \\ f_\alpha' &= \sum_{i=1}^{m} \left(\frac{a_{i\alpha}^1 sf_1 h_i}{\widehat{k}_{v_i}^{r_1}} + \frac{a_{i\alpha}^2 sf_2 h_i}{\widehat{k}_{v_i}^{r_2}} \right). \end{aligned} \tag{6}$$

(iv) *Step 4*. The energy allocated to the user via r_2 relationship further spreads to the items. Then, the energy f_α'' obtained by the item o_α can be depicted as

$$f_\alpha'' = \sum_{j=1}^{m} \frac{a_{j\alpha}^1 h_j'}{\widehat{k}_{v_j}^{r_1}}. \tag{7}$$

Thus, the total energy g_α obtained by item o_α can be defined as

$$g_\alpha = f_\alpha' + f_\alpha''. \tag{8}$$

The element $w_{\alpha\beta}^M$ of the state transition matrix W^M satisfies

$$\begin{aligned} w_{\alpha\beta}^{SM} &= \frac{1}{\widehat{k}_{v_\beta}^{r_1}} \sum_{i=1}^{m} a_{i\beta}^1 \left(\frac{a_{i\alpha}^1 sf_1}{\widehat{k}_{v_i}^{r_1}} + \frac{a_{i\alpha}^2 sf_2}{\widehat{k}_{v_i}^{r_2}} \right) \\ &+ \frac{1}{\widehat{k}_{v_\beta}^{r_1}} \sum_{j=1}^{m} \frac{a_{j\alpha}^1}{\widehat{k}_{v_j}^{r_1}} \sum_{i=1}^{m} a_{i\beta}^1 \left(\frac{a_{ij}^1 sf_1}{\widehat{k}_{v_i}^{r_1}} + \frac{a_{ij}^2 sf_2}{\widehat{k}_{v_i}^{r_2}} \right). \end{aligned} \tag{9}$$

Next, the unselected items are sorted by the final energy, producing a recommendation list.

The diffusion process of the SMD algorithm with one type of relationship (SMD1) is explained by the example in Figure 3, where $p = 0.5$.

In real social networks, there are usually multiple relationships, each of which has its unique impact on recommendation results. The multiple relationships could be introduced to the MD algorithm through the loading operation of the MSCCN model [20]. Here, another type of relationship denoted as r_3 is imported to the network of Figure 3. Suppose $sf_1 = p$ and $sf_2 = (1-p)(1-q)$. Then, $sf_3 = (1-p)q$, $p\epsilon(0,1)$, and $q\epsilon(0,1)$. In this case, the element $w_{\alpha\beta}^M$ of state transition matrix satisfies

$$\begin{aligned} w_{\alpha\beta}^{SM} &= \frac{1}{\widehat{k}_{v_\beta}^{r_1}} \sum_{i=1}^{m} a_{i\beta}^1 \left(\frac{a_{i\alpha}^1 sf_1}{\widehat{k}_{v_i}^{r_1}} + \frac{a_{i\alpha}^2 sf_2}{\widehat{k}_{v_i}^{r_2}} + \frac{a_{i\alpha}^3 sf_3}{\widehat{k}_{v_i}^{r_3}} \right) \\ &+ \frac{1}{\widehat{k}_{v_\beta}^{r_1}} \sum_{j=1}^{m} \frac{a_{j\alpha}^1}{\widehat{k}_{v_j}^{r_1}} \sum_{i=1}^{m} a_{i\beta}^1 \left(\frac{a_{ij}^1 sf_1}{\widehat{k}_{v_i}^{r_1}} + \frac{a_{ij}^2 sf_2}{\widehat{k}_{v_i}^{r_2}} + \frac{a_{ij}^3 sf_3}{\widehat{k}_{v_i}^{r_3}} \right). \end{aligned} \tag{10}$$

Figure 4 shows the process of the SMD algorithm with two types of relationships (SMD2), where $p = 0.5$ and $q = 0.5$.

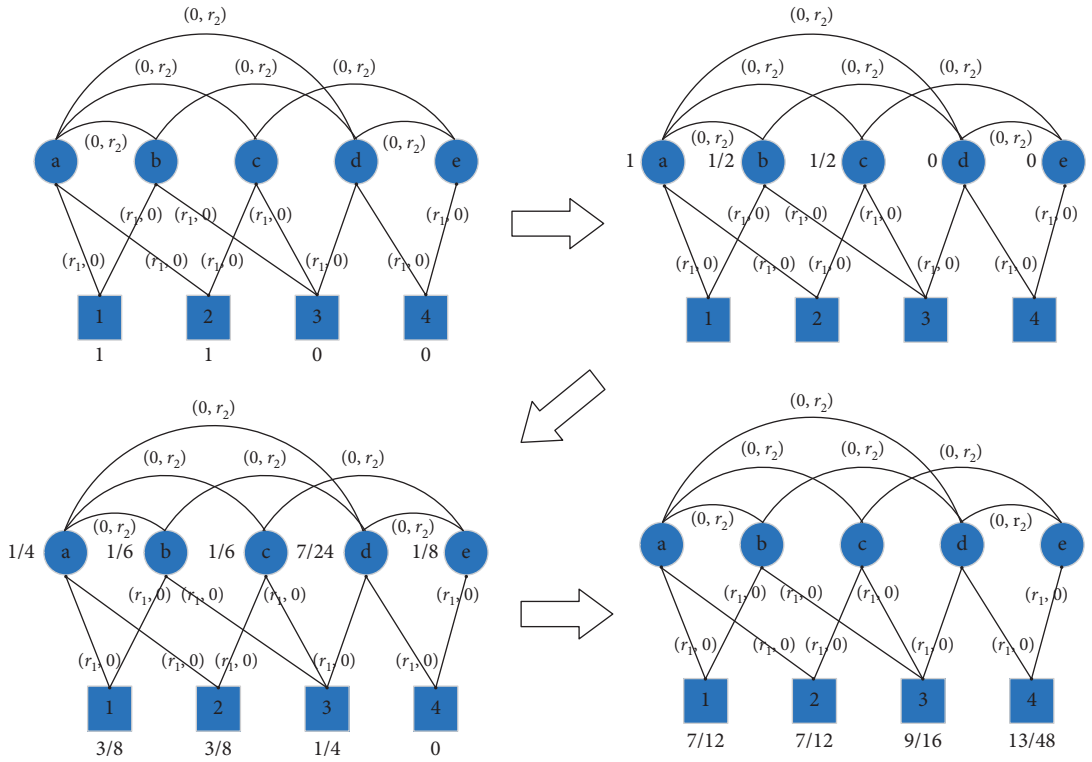


Figure 3: The diffusion process of SDM1 algorithm.

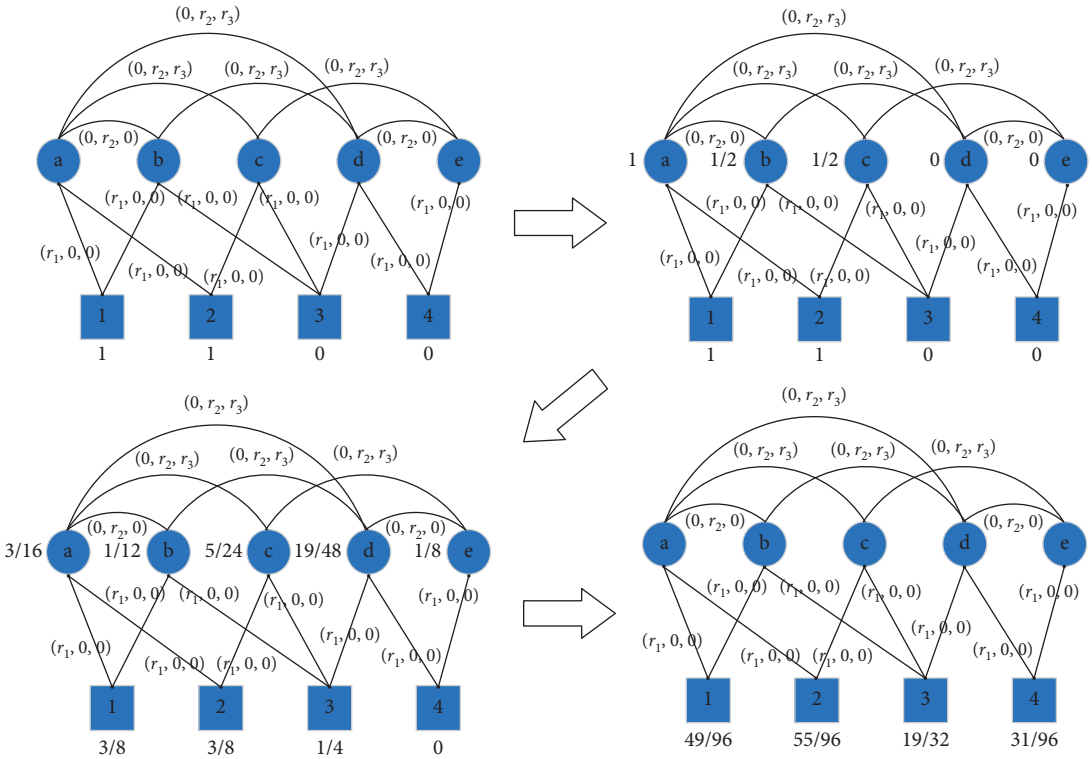


Figure 4: The diffusion process of SDM2 algorithm.

If $q = 0$ or $q = 1$, that is, $sf_2 = 0$ or $sf_3 = 0$, the SMD2 algorithm degenerates into the SMD1 algorithm.

4. Results and Discussion

4.1. Experimental Data. Two public online datasets, Epinions and FilmTrust, were selected to evaluate the performance of the proposed algorithm. The Epinions dataset contains 40,272 users, 139,738 items, 487,182 user relationships, and more than 664,000 item scores rated by users. The FilmTrust dataset provides 1,050 users, 2,071 items, >1,800 user relationships, and 35,497 item scores rated by users. The original data were preprocessed by removing the relationships of the users that have not selected any items.

4.2. Evaluation Indices. Fivefold cross validation was employed to test the performance of the SMD algorithm. Specifically, the preprocessed data were randomly split into five subsets. For each experiment, a random subset was chosen as the test set, and the remaining subsets were chosen as training sets. Five experiments were conducted to ensure that the SMD algorithm is tested on each subset. The results of the five experiments were averaged to obtain the final result.

Then, the mean ranking score (RS) [15] was introduced to evaluate the ranking accuracy of the SMD algorithm, and the Hamming distance [24] was adopted to measure the diversity of recommendation results.

Suppose the target user u_i has chosen the item o_j in the test set. Then, the ranking r_{ij} of item o_j in the recommendation list can be calculated. For the mean RS of all the items chosen by the user from the test set, the higher the value of mean RS, the better the accuracy of the recommendation algorithm. The RS of user u_i can be calculated by

$$\mathrm{RS}_i = \frac{1}{\left|E_i^P\right|} \sum_{(i\alpha)\in E_i^P} \mathrm{RS}_{i\alpha}, \tag{11}$$

where E_i^P is the number of items preferred by user u_i in test set; $i\alpha$ is the item α preferred by user i in the test set.

The Hamming distance can be defined as

$$H_{ut}(L) = 1 - \frac{Q_{ut}(L)}{L}, \tag{12}$$

where u and t are two users; $Q_{ut}(L)$ is the number of overlapping items in the recommendation lists of the two users; L is the length of the recommendation list. If $H_{ut}(L) = 1$, the two recommendation lists have no overlapping items; if $H_{ut}(L) = 0$, the two recommendation lists are identical.

4.3. Results Analysis. The numerical simulation was performed to determine the values of p and q. Figure 5 presents the simulation results of SMD1 on the two datasets with $q = 0$ and p changing between various values.

As shown in Figure 5, the optimal RS value was reached at $p = 0.4$ on Epinions dataset, indicating that SMD1 has the highest accuracy at $p = 0.4$. Similarly, it can be inferred that SMD1 has the highest accuracy at $p = 0.9$ on FilmTrust dataset.

Let O_u and O_v be the sets of items chosen by users u and v, respectively. The greater the number of overlapping items in the two sets is, the more likely that the two users have the same interests, and the greater they influence each other. The user similarity can be defined as

$$f_{uv} = \frac{O_u \cap O_v}{O_u \cup O_v}. \tag{13}$$

If $f_{uv} > 0.2$, the two users have similar interests.

Figure 6 displays the simulation results of SMD2 with both p and q changing between various values.

As shown in Figure 6, the RS value reached the minimum at $p = 0.4$ and $q = 0.2$ on Epinions dataset. This means, on Epinions dataset, the SMD2 has the highest accuracy at $p = 0.4$ and $q = 0.2$. Similarly, the SMD2 has the highest accuracy at $p = 0.9$ and $q = 0.9$ on FilmTrust dataset. Since r_3 relationship is denser than r_2 relationship, the recommendation accuracy increases as q approaches 1 on FilmTrust dataset.

For comparison, traditional recommendation algorithms, namely, HeatS algorithm and Hybrid algorithm [25], were also applied on Epinions dataset and FilmTrust dataset. The RS values of the Hybrid algorithm were minimized on the two datasets, when the regularization parameter λ was 0.67 and 0.5, respectively. Table 1 compares the results of the contrastive algorithms at the optimal values of p and q.

As shown in Table 1, SMD2 achieved slightly higher recommendation accuracy than SMD1. The accuracy of the two algorithms was marked higher than that of the traditional algorithms on the two datasets. The results indicate that the recommendation system becomes more accurate, thanks to the introduction of multiple social relationships.

The diversity of the recommendation list is also an important indicator of recommendation quality. Figure 7 shows the diversity trends of recommendation list on Epinions dataset and FilmTrust dataset, with the list length changing between various values.

As shown in Figure 7, the length of the recommendation list is positively correlated with the similarity between recommendation lists for different users. On the two datasets, SMD2 and SMD1 had lower recommendation diversity than the MD algorithm. The relatively low diversity is the result of the following: After user relationship(s) is introduced, the recommendation list for a user is affected by his/her historical scores and also the historical scores of the relevant users. In other words, the recommendation list of a user will bear a high resemblance to that of users, who has social relationships with the user. The HeatS algorithm achieved higher recommendation diversity than the MD algorithm because it attracts the historical scores of the user to the less popular items.

In the real world, some users only select very few items. For implicitly, the users who have purchased fewer than 30

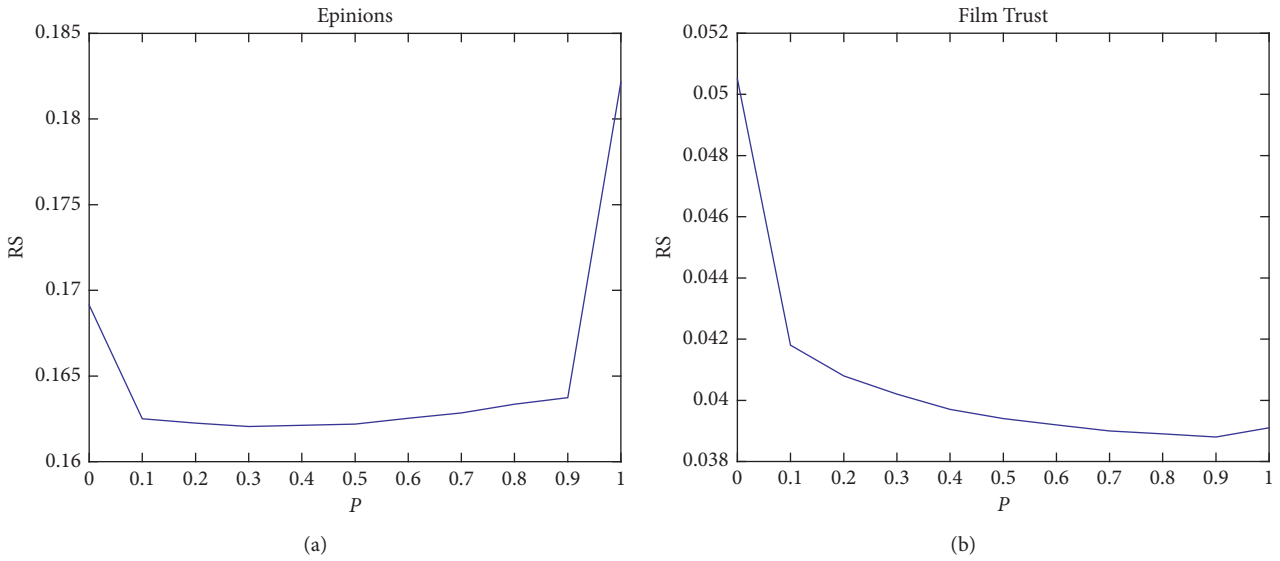


Figure 5: The simulation results of SMD1. (a) Epinions dataset. (b) FilmTrust dataset.

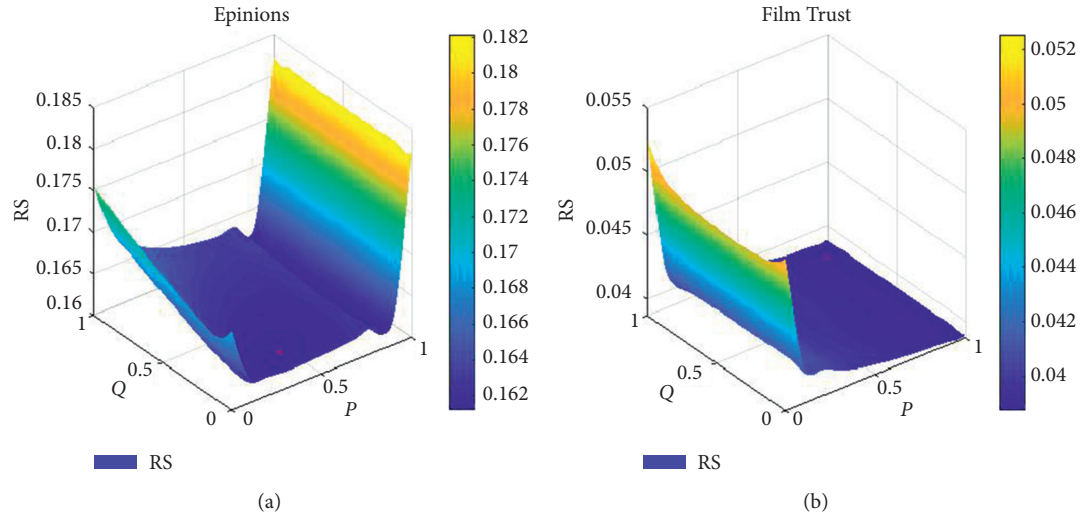


Figure 6: The simulation results of SMD2. (a) Epinions dataset. (b) FilmTrust dataset.

Table 1: The results of the contrastive algorithms.

Dataset	MD	HeatS	Hybrid	SMD1	SMD2
Epinions	0.18216	0.20215	0.17063	0.16203	0.16128
FilmTrust	0.04032	0.04232	0.04015	0.03894	0.03891

items were defined as small-degree users, and the other users as generous-degree users. The small-degree users generate relatively few data to be used for recommendation. Therefore, the recommendation to such users is usually of low accuracy. Then, the users with $k_i \leq 30$ (k_i is the degree of user u_i) on the two selected datasets were treated as a small-degree user group. Figure 8 shows the recommendation results of SMD for this user group on the two datasets, with $q = 0$ and p changing between various values.

As shown in Figure 8, SMD1 algorithm achieved the minimum RS at $p = 0.2$ and $p = 0.6$, respectively, on Epinions and FilmTrust datasets. The p value decreased from the scenario in Figure 5, where SMD1 achieved the highest accuracy at $p = 0.4$ and $p = 0.9$, respectively. The p value determines the number of resources that need to be transmitted through the social network to reach the item during the diffusion process. The smaller the p value, the greater the resource demand.

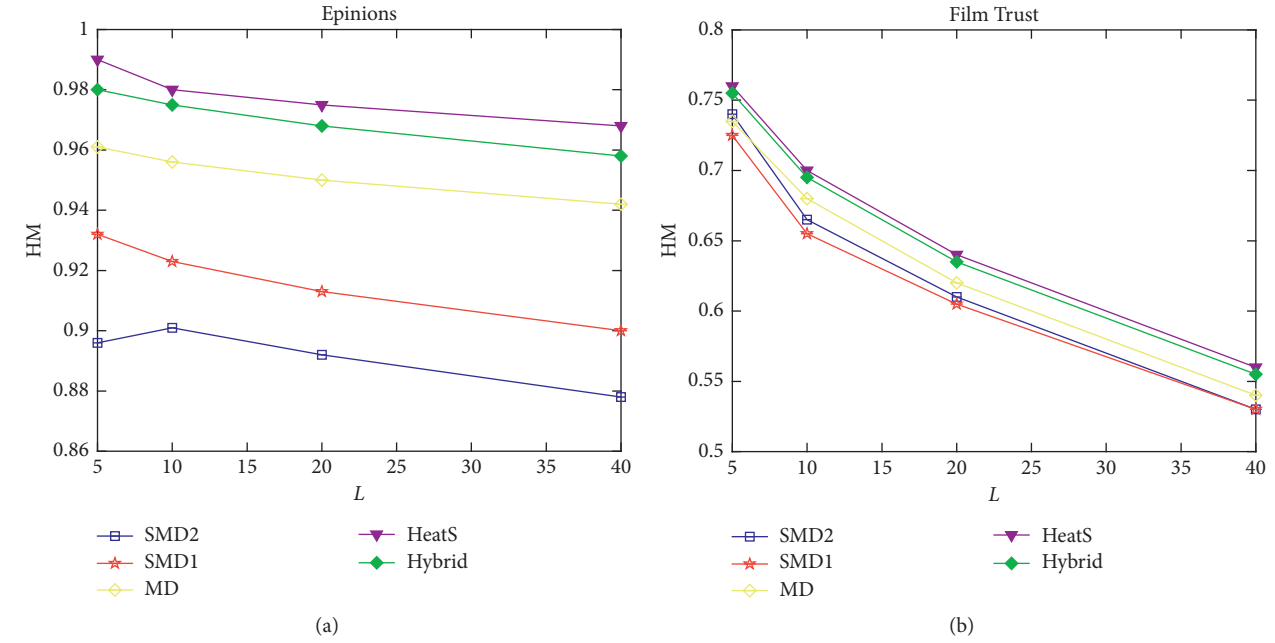


Figure 7: The diversity of various algorithms on two datasets, (a) Epinions dataset. (b) FilmTrust dataset.

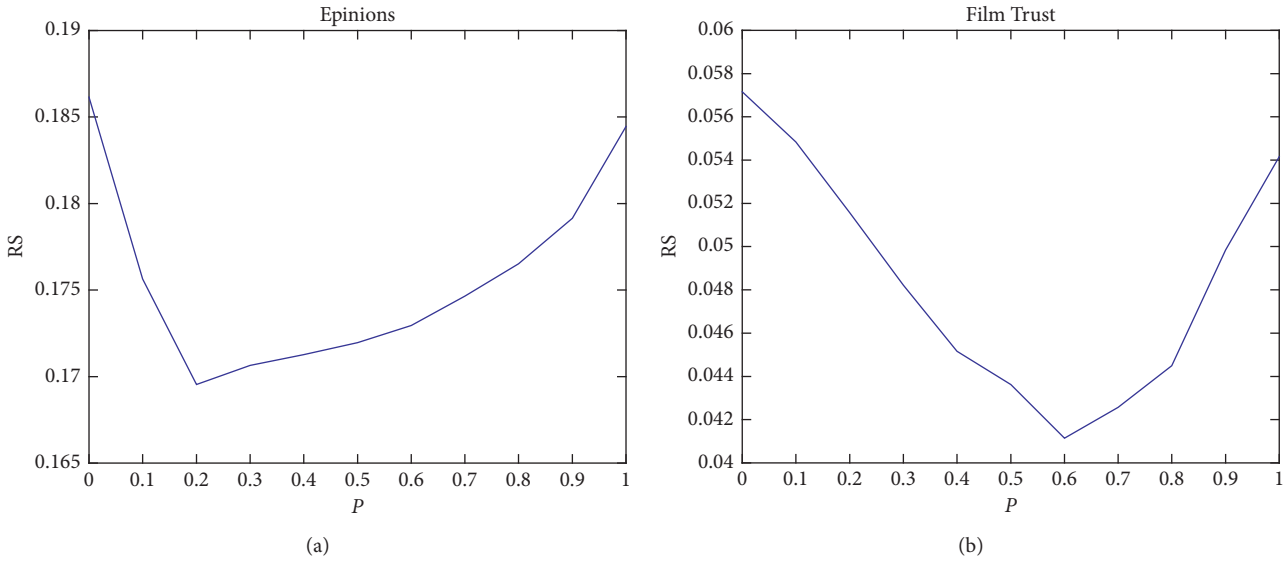


Figure 8: The performance of SMD on small-degree users on the two datasets. (a) Epinions dataset. (b) FilmTrust dataset.

To further explore the relationship between parameters p and k_i, the RS values were grouped by user degree. Figure 9 shows the p values corresponding to the optimal RS values on the two datasets.

As shown in Figure 9, the relationships between p and k_i exhibited a consistent trend on the two datasets: the smaller the degree of the target user, the smaller the p value for the RS to reach the optimal value. A user with a small degree chooses only a few items and provides limited information for the recommendation algorithm. Therefore, the recommendation accuracy for small-degree users needs to be improved based on social network information.

To demonstrate the superiority of SMD in the recommendation for small-degree users, the small-degree users satisfying $k_i \leq 30$ were grouped by k_i, as p and q reached the optimal values on the two datasets. Figure 10 compares the optimal RS values under SMD and MD for small-degree users.

As shown in Figure 10, the highest RS values were achieved by MD algorithm on Epinions dataset, followed in

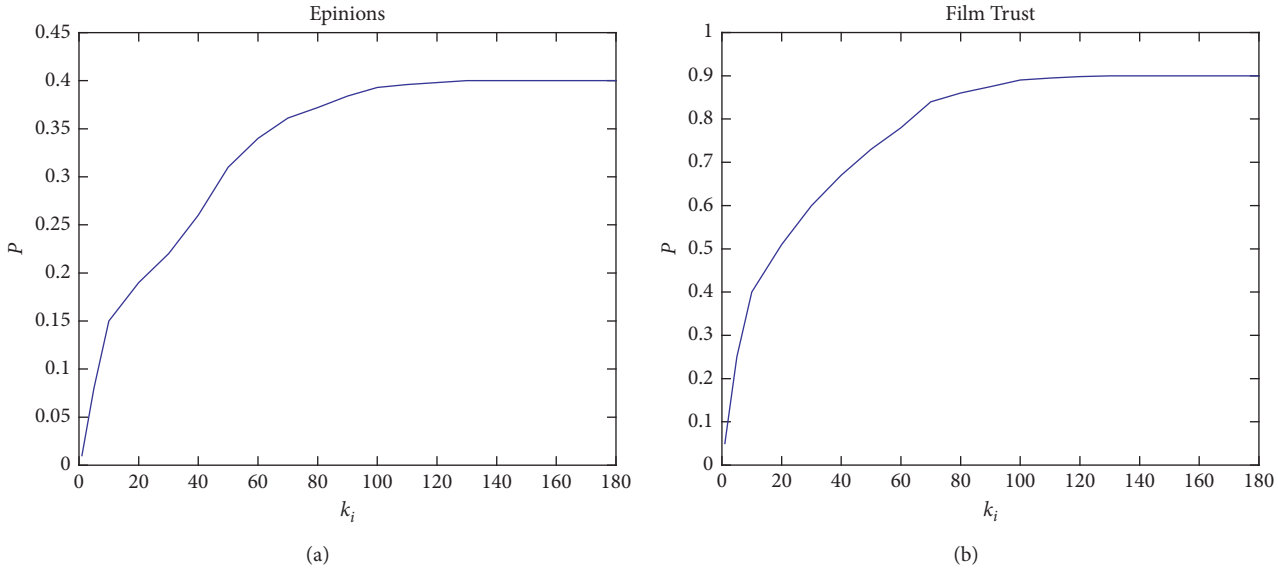


Figure 9: The optimal p value as a function of item degree k_i. (a) Epinions dataset. (b) FilmTrust dataset.

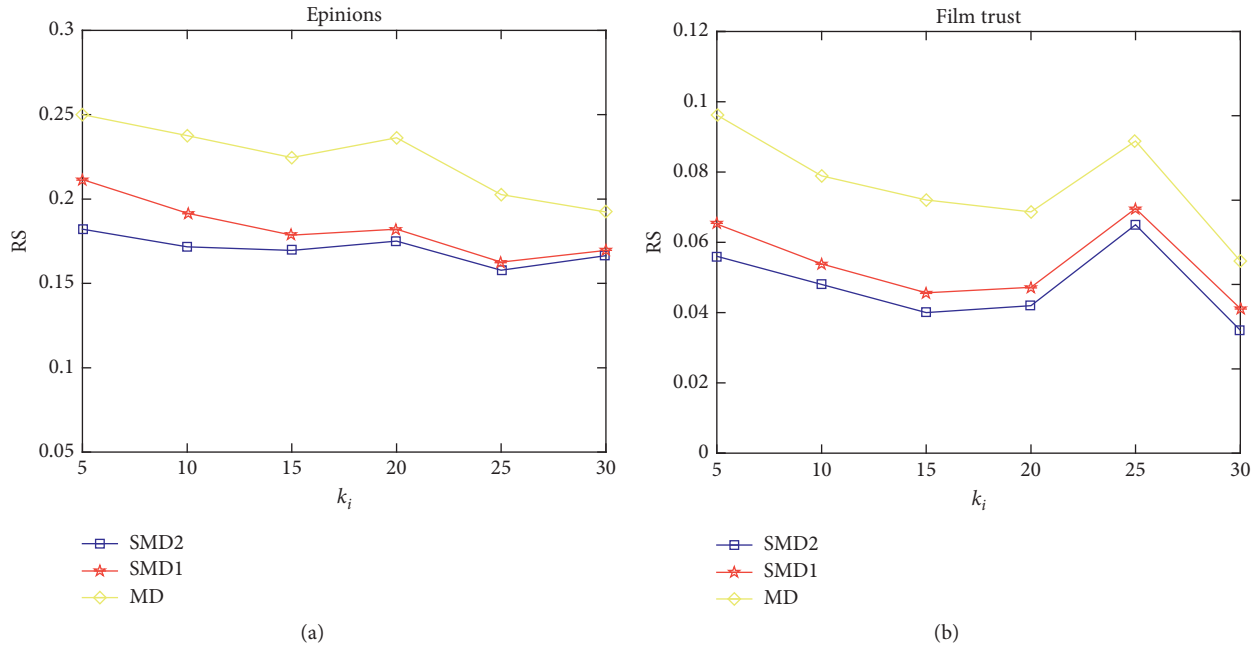


Figure 10: The comparison of optimal RS values under SMD and MD for small-degree users. (a) Epinions dataset. (b) FilmTrust dataset.

turn by SMD1 and SMD2. That is, for small-degree users, SMD outperforms the MD in recommendation accuracy; the advantage increases with the reduction of user degree. The same was observed on FilmTrust dataset. To sum up, SMD is more effective than MD in the recommendation for small-degree users.

5. Conclusions

This paper introduces social information to the MSCCN model to create a complex network of multiple relationships. On this basis, the MD algorithm was improved into the SMD. Experimental results show that the SMD made a highly accurate recommendation to users because it considers the potentially useful items of different users in the social network. It is also learned that SMD2 outperformed SMD1 and MD, indicating that the integration of two types of relationships is better than incorporating only one type of relationship or considering no social relationship. In other words, the recommendation accuracy can be greatly improved by introducing multiple social relationships between users. In addition, the experiments also demonstrated that

the SMD is highly effective in making accurate recommendations to small-degree users. The future research will discover the implicit social relationships between users and evaluate their impacts on recommendation results.

Data Availability

The basic data used in this paper were downloaded from two public datasets online: the Epinions http://www.trustlet.org/epinions.html and the FilmTrust https://www.librec.net/datasets.html.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was funded by Shandong Provincial Natural Science Foundation, China, Grant no. ZR2017MG011.

References

[1] D. Bawden, C. Holtham, and N. Courtney, "Perspectives on information overload," *Aslib Proceedings*, vol. 51, no. 8, pp. 249–255, 2013.

[2] Y. Liu, H. Peng, and J. Wang, "Verifiable diversity ranking search over encrypted outsourced data," *Computers. Materials & Continua*, vol. 55, no. 1, p. 37, 2018.

[3] S. Fekri-Ershad, "Gender classification in human face images for smart phone applications based on local texture information and evaluated Kullback-Leibler divergence," *Traitement Du Signal*, vol. 36, no. 6, pp. 507–514, 2019.

[4] R. Meng, S. G. Rice, J. Wang, and X. Sun, "A fusion steganographic algorithm based on faster *R*-CNN," *Computers. Materials & Continua*, vol. 55, no. 1, pp. 1–16, 2018.

[5] B. K. Lee and W. N. Lee, "The effect of information overload on consumer choice quality in an on-line environment," *Psychology & Marketing*, vol. 21, no. 3, pp. 159–183, 2010.

[6] L. Lü, M. Medo, C. H. Yeung, Y.-C. Zhang, Z.-K. Zhang, and T. Zhou, "Recommender systems," *Physics Reports*, vol. 519, no. 1, pp. 1–49, 2012.

[7] Y. Li, "Design and implementation of intelligent travel recommendation system based on internet of things," *Ingénierie Des Systèmes D'Information*, vol. 23, no. 5, pp. 159–173, 2018.

[8] N. Hasanzadeh and Y. Forghani, "Improving the accuracy of *M*-distance based nearest neighbor recommendation system by using ratings variance," *Ingénierie Des Systèmes D Information*, vol. 24, no. 2, pp. 131–137, 2019.

[9] J. B. Schafer, D. Frankowski, J. Herlocker et al., "Collaborative filtering recommender system," in *The Adaptive Web, Lecture Notes in Computer Science*, vol. 4321, pp. 291–324, Springer-Verlag, Berlin, Germany, 2007.

[10] M. J. Pazzani and D. Billsus, "Content-based recommendation systems," in *The Adaptive Web, Lecture Notes in Computer Science*, vol. 4321, pp. 325–341, Springer-Verlag, Berlin, Germany, 2007.

[11] S. Maslov and Y.-C. Zhang, "Extracting hidden information from knowledge networks," *Physical Review Letters*, vol. 87, no. 24, pp. 248701–248705, 2001.

[12] K. Goldberg, T. Roeder, D. Gupta, and C. Perkins, "Eigentaste: a constant time collaborative filtering algorithm," *Information Retrieval*, vol. 4, no. 2, pp. 133–151, 2001.

[13] M. J. Pazzani, "A framework for collaborative. content-based and demographic filtering," *Artificial Intelligence Review*, vol. 13, no. 5/6, pp. 393–408, 1999.

[14] K. Yoshii, M. Goto, K. Komatani, T. Ogata, and H. G. Okuno, "An efficient hybrid music recommender system using an incrementally trainable probabilistic generative model," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 2, pp. 435–447, 2008.

[15] S. Bin, G. Sun, N. Cao et al., "Collaborative filtering recommendation algorithm based on multi-relationship social network," *Computers, Materials & Continua*, vol. 60, no. 2, pp. 659–674, 2019.

[16] J. M.. Hu and X. Lin, "Design and implementation of recommendation algorithm based on user-socialized resource-vocabulary three-part graph," *Information Studies: Theory & Application*, vol. 39, no. 3, pp. 130–134, 2016.

[17] W. Li, J. Qi, Z. Yu, and D. Li, "A social recommendation method based on trust propagation and singular value decomposition," *Journal of Intelligent & Fuzzy Systems*, vol. 32, no. 1, pp. 807–816, 2017.

[18] B. Beirami and M. Mokhtarzade, "Spatial-spectral random patches network for classification of hyperspectral images," *Traitement Du Signal*, vol. 36, no. 5, pp. 399–406, 2019.

[19] W. Meng, C. Mao, J. Zhang, J. Wen, and D. Wu, "A fast recognition algorithm of online social network images based on deep learning," *Traitement Du Signal*, vol. 36, no. 6, pp. 575–580, 2019.

[20] G. Sun and S. Bin, "A new opinion leaders detecting algorithm in multi-relationship online social networks," *Multimedia Tools and Applications*, vol. 77, no. 4, pp. 4295–4307, 2018.

[21] G. Sun and S. Bin, "Router-level internet topology evolution model based on multi-subnet composited complex network model," *Journal of Internet Technology*, vol. 18, no. 6, pp. 1275–1283, 2017.

[22] S. Bin and G. Sun, "Optimal energy resources allocation method of wireless sensor networks for intelligent railway systems," *Sensors*, vol. 20, no. 2, p. 482, 2020.

[23] G. Sun, S. Bin, M. Jiang et al., "Research on public opinion propagation model in social network based on blockchain," *Computers, Materials & Continua*, vol. 60, no. 3, pp. 1015–1027, 2019.

[24] A. Bellogín, I. P. Castells, and I. Cantador, "Statistical biases in information retrieval metrics for recommender systems," *Information Retrieval Journal*, vol. 20, no. 6, pp. 606–634, 2017.

[25] T. Zhou, Z. Kuscsik, J.-G. Liu, M. Medo, J. R. Wakeling, and Y.-C. Zhang, "Solving the apparent diversity-accuracy dilemma of recommender systems," *Proceedings of the National Academy of Sciences*, vol. 107, no. 10, pp. 4511–4515, 2010.

Hindawi
Discrete Dynamics in Nature and Society
Volume 2020, Article ID 1041927, 8 pages
https://doi.org/10.1155/2020/1041927

Research Article

A GIS Based Unsteady Network Model and System Applications for Intelligent Mine Ventilation

Hui. Liu, Shanjun Mao, Mei. Li, and Pingyang Lyu

Institute of Remote Sensing and Geographic Information System, Peking University, Beijing 100871, China

Correspondence should be addressed to Shanjun Mao; sjmao_@pku.edu.cn

Received 26 June 2020; Revised 14 September 2020; Accepted 12 October 2020; Published 28 October 2020

Academic Editor: Chi-Hua Chen

With the development of state-of-the-art technology, such as the artificial intelligence and the Internet of Things, the construction of "intelligent mine" is being vigorously promoted, where the intelligent mine ventilation is one of the primary concerns that provides the efficient guarantee for safety production in the underground coal mine system. This study aims to integrate the geographical information system (GIS) and the unsteady ventilation network model together, to provide location based information and online real-time support for the decision-making system. Firstly, a GIS based unsteady network model is proposed, and its algorithm steps are brought out. Secondly, a prototype web system, named 3D VentCloud, is designed and developed based on the front and end technique, which effectively integrates the proposed algorithms. Thirdly, the model is validated, and the system is applied to a real coal mine for ventilation solution, which demonstrates that the model is reasonable and practical. The online simulation system is efficient in providing real-time support. The study is potential and is expected to guide the real-time coal mine safety production.

1. Introduction

The "intelligent mine" is being constructed and developed vigorously in recent years, which is expected to provide reliable, fast, smart and accurate decision support for safety operation and production in underground coal mine. Mine ventilation system is one of the most important and basic systems to ensure safety production. An efficient ventilation system is potential to guarantee a health working environment for coal miners and ensure sustainable development of the underground coal mine system. The concept of the intelligent mine ventilation is being proposed, while there are still a lot of scientific issues to be solved [1, 2].

The current research on ventilation network model and ventilation software is mainly focused on the traditional steady calculation model. However, mine ventilation system essentially belongs to a dynamic complex system with geospatial and temporal characteristics. For example, although the ventilation state is steady during normal production of coal mine, various internal disturbances and ground air pressure would have certain influence on airflow and pressure inside the laneway; thus, it causes airflow pulsation [3]. Besides, due to gas outburst, mine fire, and other major accidents, the ventilation state will change instantaneously and cause transient disorder of airflow inside the laneway. Therefore, the ventilation states during normal production or disaster should all be regarded as unsteady flow; thus, it requires to establish a unified ventilation simulation model, and that is well adapted with the concept of the intelligent mine ventilation [4].

Besides, the current data acquisition and processing procedures for mine ventilation lack an online integrated system to better adapt to intelligent mine ventilation system. Nowadays, there are a lot of ventilation software systems that implemented informatization for mine ventilation, for example, Polish Academy of Sciences developed VENTGRAPH and Mine Fire Simulation software [5–7]. Chasm company in Australia developed Ventsim software, which is widely used around the world [8]. The US and UK also developed some ventilation software systems such as VentPC 2003, Ventilation Design, VR-MNE, and Datamine [9]. There are also some successful domestic ventilation

systems such as MFire, MVSS, and VentAnaly developed by Chinese institutes and universities [10, 11]. With the deep understanding for mine ventilation system and the great demand for practical production, mine ventilation belongs to a typical geographical environment, which has spatial, temporal, and attribute data as well as the spatial topological relationship [12]. Fortunately, the geographic information system (GIS) is regarded as one of the most effective ways to manage the geospatial data [13]. Therefore, based on traditional GIS technology, a great number of ventilation software systems have been developed. For instance, LongruanGIS, LKGIS, and VRMine GIS platform and some secondary developed systems based on AutoCAD or ArcGIS are developed and have been widely used [14–17].

However, it is difficult to deal with the spatial, temporal, and attribute data as well as the topological relationship using AutoCAD. In addition, the ventilation software systems based on GIS are mostly focusing on the combination of traditional GIS or static GIS with steady ventilation simulation model. Thus, a unified unsteady ventilation model integrated with GIS technology is needed to provide online and fast decision support to better satisfy the requirement of intelligent mine ventilation.

Therefore, this study aims to provide a fast online ventilation simulation solution to promote the construction of intelligent mine ventilation by integrating the unsteady ventilation network model and GIS as well as the system development. Firstly, the basic ventilation theory is introduced and an algorithm of GIS based unsteady network model is proposed. Secondly, based on the data structure of the ventilation model as well as the laneway network, a prototype simulation web system, called 3D VentCloud, is designed and developed. Thirdly, two experiments are implemented to validate the ventilation model and apply the web system, respectively.

2. GIS-Based Unsteady Network Model

Mine ventilation is a daily work to ensure safety production of coal mine, which requires accurate simulation to learn about the onsite ventilation situation. Rather than the traditional steady ventilation simulation model, this section mainly introduces the unsteady network model based on GIS technology as well as the model integration with the online web system.

2.1. Mathematical Models. Firstly, an ideal geometric laneway model is established to build the basic equation for the unsteady ventilation network model. As can be seen in Figure 1, the laneway length is L and cross section area is A, without considering the compressible properties of airflow.

Without any external force, the airflow inside the laneway belongs to the steady state. If the wind pressure at the ends of the laneway suddenly changes, there will be an equivalent external force F to the ends of the air column. According to the Newton's second law, suppose that the external force gives the acceleration a, and the air column has the corresponding equations [18]:

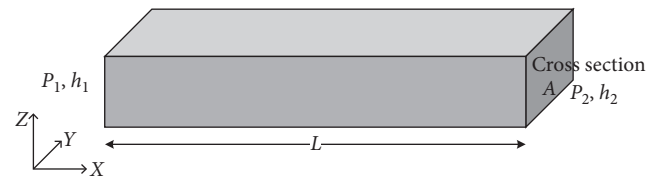


Figure 1: An ideal geometric laneway model.

$$a = \frac{dv}{dt} = \frac{1}{A}\frac{dQ}{dt} = \frac{F}{m} = \frac{\Delta PA}{AL\rho} = \frac{\Delta P}{L\rho}, \tag{1}$$

$$\Delta P = (p1 - p2) + \left(\frac{1}{2}\rho v_1^2 - \frac{1}{2}\rho v_2^2\right) + \rho g \Delta h + h_r - RQ^2,$$

$$\frac{dQ}{dt} = \frac{A}{L\rho}\left[(p1 - p2) + \left(\frac{1}{2}\rho v_1^2 - \frac{1}{2}\rho v_2^2\right) + \rho g \Delta h + h_r - RQ^2\right], \tag{2}$$

where R is the ventilating resistance, p_1 and p_2 are static pressures, v_1 and v_2 are the velocities at the laneway ends, and h_r is the external power. Assuming that a laneway is single and horizontal, both ends are connected with the atmosphere. When a fan placed at the end of the laneway suddenly starts, the pressure difference in equation (2) is 0, then we obtain

$$\frac{dQ}{dt} = \frac{A}{L\rho}\left[h_r - RQ^2\right]. \tag{3}$$

As for the airflow inside the laneway network, the air is assumed to follow three basic laws of wind flow, which are the law of ventilation resistance, the law of pressure balance in loops, and the law of wind balance at nodes. These three laws represent the restriction and equilibrium relationships for three basic ventilation parameters, that is, air volume and ventilating resistance for laneway branch and node pressure in the ventilation network model [19]. A simple ventilation network graph is shown in Figure 2.

The three basic laws of wind flow can be described as

$$h_i = R_i \cdot |Q_i| \cdot Q_i, \quad (i = 1, 2, \ldots, N),$$

$$\sum_{i=1}^{N} \varepsilon_{ki} Q_i = 0, \quad k = 1, 2, 3, \ldots, (b-1), \tag{4}$$

$$\sum_{i=1}^{N} C_{ji}\beta_i \frac{dQ}{d\tau} = \sum_{i=1}^{N} C_{ji}\left(\Delta P_i - R_i Q_i |Q_i|\right), \quad j = 1, 2, 3, \ldots, M, \tag{5}$$

where N is the number of the laneway branches, and M is the number of network nodes. ε_{ki} and C_{ij} are the elements from basic correlation matrix and independent loop matrix of the network graph, and the value is 0, 1 or −1. $\Delta P_i = h_{Fj}(Q) \mp h_{Nj}(Q)$ stands for the ventilation power, including fan power $h_{Fj}(Q)$ and natural power $h_{Nj}(Q)$. Equation (5) can also be simplified to the matrix form of loop equations:

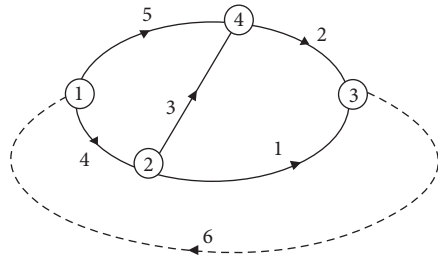


Figure 2: Ventilation network graph.

$$\frac{dQ(\tau)}{d\tau} = C^T\left(C\beta C^T\right)^{-1} D, \tag{6}$$

where C is the loop matrix, and the elements in D are $D_j = \Delta P_j - \sum C_{ji} R_i Q_i(\tau)|Q_i(\tau)|$, $j = 1, 2, \ldots, b$. Based on this, the Runge-Kutta iterative algorithm can be applied to simulate the air volume changing with time for each laneway branch [20, 21]. The following is our specific solution based on the classical fourth-order Runge-Kutta method.

$$Q(\tau_{k+1}) = Q(\tau_k) + \frac{h}{6}(K_1 + 2K_2 + 2K_3 + K_4), \tag{7}$$

$$K_1 = f(Q(\tau_k)), \tag{8}$$

$$K_2 = f\frac{Q(\tau_k) + hK_1}{2}, \tag{9}$$

$$K_3 = f\frac{Q(\tau_k) + hK_2}{2}, \tag{10}$$

$$K_4 = f\frac{Q(\tau_k) + hK_3}{2}. \tag{11}$$

2.2. Data Models for Ventilation Network. Based on GIS theories and methods, the data model for unsteady ventilation network model is designed. Firstly, the ventilation network graph is described as $G = (V, E)$, where $V = \{v_1, v_2, \ldots, v_m\}$ is a set of nodes in network graph, and $E = \{e_1, e_2, \ldots, e_n\}$ is a set of laneway branches in network graph. Consider that airflow on every laneway branch has a wind direction; thus, the ventilation network graph is a directed graph. Then, we obtained the adjacency matrix, incidence matrix, cut set matrix, loop matrix, and path matrix of the network graph, which stores the topological information of the ventilation network graph. Besides, based on the perspective of GIS, the ventilation network graph and model belong to typical geospatial objects, including spatial data, attribute data, temporal data, and the spatial relationship, where the spatial relationship mainly includes the topological relation of the ventilation network graph, which connected the laneway branches and nodes, and is stored as point-line (node-branch) indexed structure based on GIS data organization, as can be seen in Table 1.

The specific data structure for laneway branch and nodes is established and presented in Table 2.

Table 1: Point-line indexed structure for topological relation.

Laneway branch ID (line)	Start node ID (point)	End node ID (point)

2.3. Algorithm Design. The whole algorithm is designed for the GIS based unsteady ventilation network model. Table 3 shows the specific procedure of the whole structure. This model and data organization structure provides more details of geospatial and attribute data for ventilation network by integrating geographical information.

3. System Design and Development

A prototype system, called 3D VentCloud, is designed in this study, which is developed by combing the front-end and the back-end technologies. The front-end is developed by using Html, JavaScript, and CSS. The back-end mainly includes the model and algorithm, which is developed by C++ and Python. Besides, Tornado is adopted as the server to transmit data between the front-end and back-end, and thus the two ends are connected. The system mainly includes three layers. As presented in Figure 3, they are, respectively, technology layer, service layer, and application layer.

Each layer and its functions are specifically described as follows:

(1) Technology layer. This layer contains the data source and the GIS based unsteady ventilation network model, which is the key technology for the system development. The data source includes the original obtained centerline data of the laneway network, which is processed and stored in GIS database as GeoJson format, including the geospatial, attribute, and temporal data of the laneway branch and nodes. The GIS based unsteady network model includes the generation of the topological relationship of the ventilation network graph, and the unsteady ventilation simulation model, where the topological relation and the dynamic simulation results are linked by the point-line indexed structure. These models jointly constructed the core model library and the key technology for the ventilation simulation system.

(2) Service layer. In this layer, the Tornado architecture is applied as the service, and WebSocket interface is utilized to link the front-end of JavaScript with the

Table 2: Data structure for laneway branch and node.

Data type	Laneway branch attribute (line)	Data type	Node attribute (point)
Int	ID	Int	ID
String	name	String	name
String	Kind	Double	Coordinate (x, y, z)
Double	Length	Double	Pressure
Double	Cross section area	Double	Humidity
Double	Perimeter	Double	Density
Double	Wind drag	Double	Temperature
Double	Ventilating resistance	vector<int>	Adjacent laneway ID
Vector <double>	Air volume (time series)		

Table 3: The algorithm for GIS Based Unsteady Network Model.

Algorithm 1. Algorithm steps for GIS based unsteady network model
Step 1: Establish the ventilation network graph, initialize the geospatial and attribute data for laneways and nodes.
Step 2: Generate the topological relation for the ventilation network graph and stored as point-line indexed structure and basic incidence matrix.
Step 3: Sort the laneway branches according to wind drag, and generate the minimum spanning tree based on Dijkstra shortest path algorithm.
Step 4: Search the cotree branches, generate the independent loop and loop matrix.
Step 5: Main iteration based on Runge-Kutta algorithm; see equations (7)–(11).
Step 6: Save the results as time series air volume for each laneway branch based on point-line indexed structure.

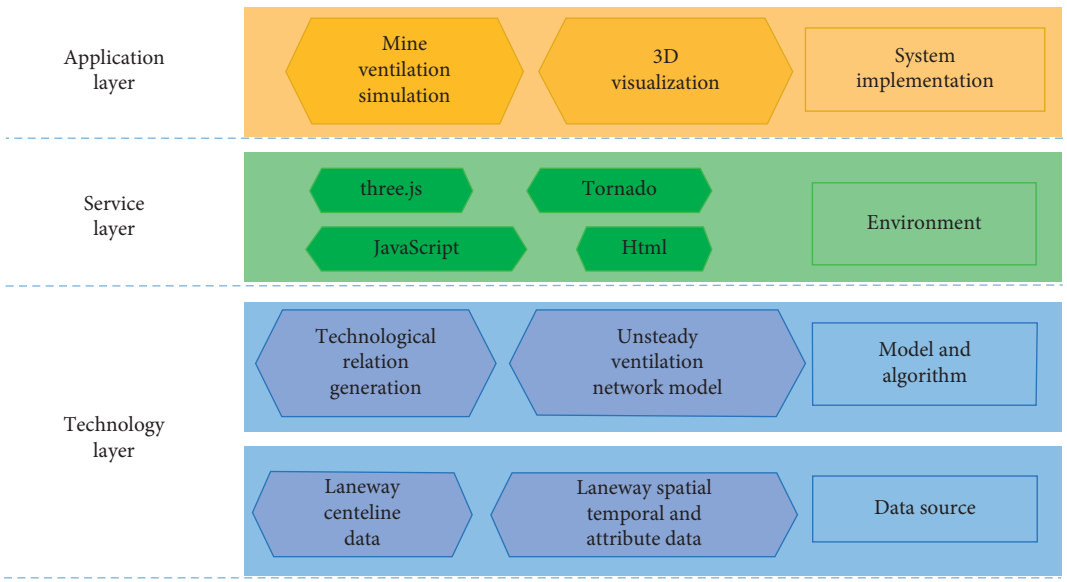


Figure 3: The architecture of 3D VentCloud system.

back-end of Python and C++ program to realize the data transmission. Three .js are adopted to implement the 3D visualization function.

(3) Application layer. This layer provides the unsteady ventilation simulation function, 3D visualization and spatiotemporal attribute query function of the laneway network, and simulation results. The system can read the original data of laneway centreline to store it and display the 3D laneway model and output the .stl file of the laneway geometric model. When the simulation function is implemented, each of the laneway branches can also be rendered with different colors based on different air volume, and the geospatial coordinates and attribute data of the laneway can be queried and displayed to show more geographical laneway details.

4. Results and Discussion

4.1. GIS-Based Unsteady Ventilation Network Model Validation. In order to verify the GIS based nonsteady network model before field application, we acquired the data from reference [3]. The simplified mine ventilation network diagram and its minimum spanning tree are shown in Figure 4. The network diagram has 11 laneway branches including one virtual branch, and 8 nodes. Table 4 presents

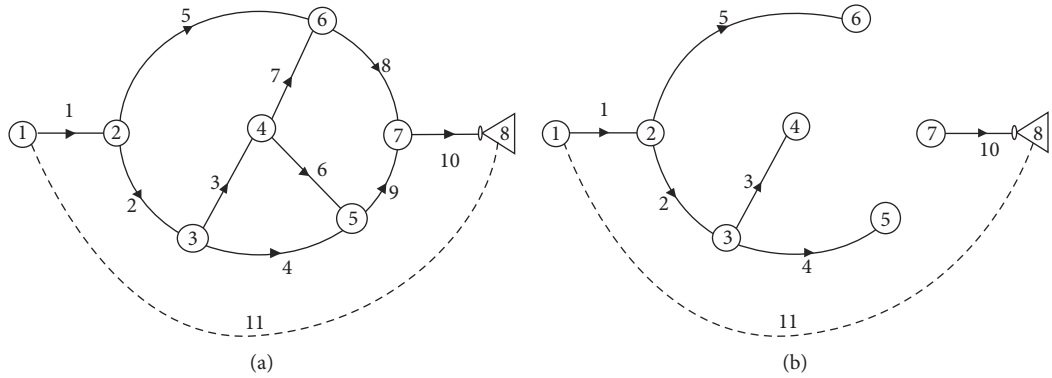


Figure 4: Mine ventilation network diagram (a) and its minimum spanning tree (b).

Table 4: Basic ventilation parameters for the network.

Laneway branch number	Start node number	End node number	Wind drag (kg·m^{-4})	Ventilating resistance (Pa)
1	1	2	0.020	105
2	2	3	0.001	0
3	3	4	0.010	0
4	3	5	0.050	0
5	2	6	0.069	0
6	4	5	0.010	0
7	4	6	0.020	40
8	6	7	0.0045	0
9	5	7	0.010	0
10	7	8	0.02	150
11	8	1	0	0

the basic ventilation parameters for the network. The case study investigated wind flow and pressure disturbance caused by the mine car movement in the transportation laneway, which belongs to the unsteady flow phenomenon.

In this case study, the mine car is moving windward inside the transportation laneway (with the branches numbers 2 and 3), which would lead to the increase of dynamic wind drag and the corresponding decrease of air volume in laneway branches 2 and 3. Based on the nonsteady network model, we simulated the variation regularity over time of air volume for each laneway, as can be seen in Figure 5. The positive value represents the reduced air, and the negative value stands for the increased value of air quantity. It is obvious that the transportation laneway with branches numbers 2 and 3 has the maximum air reduced quantity, while branch number 5 has the maximum air increased quantity, which is easy to understand because most of the airflow from the air intake laneway flows into the laneway (with branch number 5) that is connected with transportation laneway in node 2.

Besides, we also investigated the particular laneways and nodes to observe the changing regularity over time of air volume and pressure after mine car stops moving to windward. As presented in Figures 6 and 7, the laneway branches 2 and 5, nodes 3 and 5 are taken as examples to show the nonsteady characteristics of air volume and pressure change. It shows that, with the mine car moving to windward, air volume reduction in laneways 2 and 5 presents a significantly rising trend and falling trend, respectively, and then keeps a steady state. After the mine car stops moving to windward, air volume reduction in laneways 2 and 5 appears with an opposite tendency and finally reaches the initial steady state. The wind pressure variation has the similar regularity. The result is in accord with the published research in [3] and is consistent with the variation law of airflow and wind pressure in ventilation network model.

4.2. A Case Study Application for a Real Mine Ventilation Network. This study is also applied to Xinqiao coal mine (located in Henan province, China) to extend the practicality of the model and investigate the real mine ventilation to further promote the development of intelligent ventilation. Therefore, the real mine ventilation network data from Xinqiao coal mine is obtained, which contains 290 laneways and 222 nodes. The three-dimensional geometric model of Xinqiao laneway network is read and visualized by 3D VentCloud system, as presented in Figure 8.

The specific spatial data and attribute data include the ID and name of the laneway, the ID and coordinates of the start node and end node, the area and perimeter of laneway section, laneway length, laneway type, wind drag, ventilating resistance, and friction resistance. Here, some of the basic ventilation parameters are shown in Table 5.

The topological structure for ventilation network graph of Xinqiao coal mine is established based on GIS to simulate the ventilation state. At first, the shortest path algorithm is

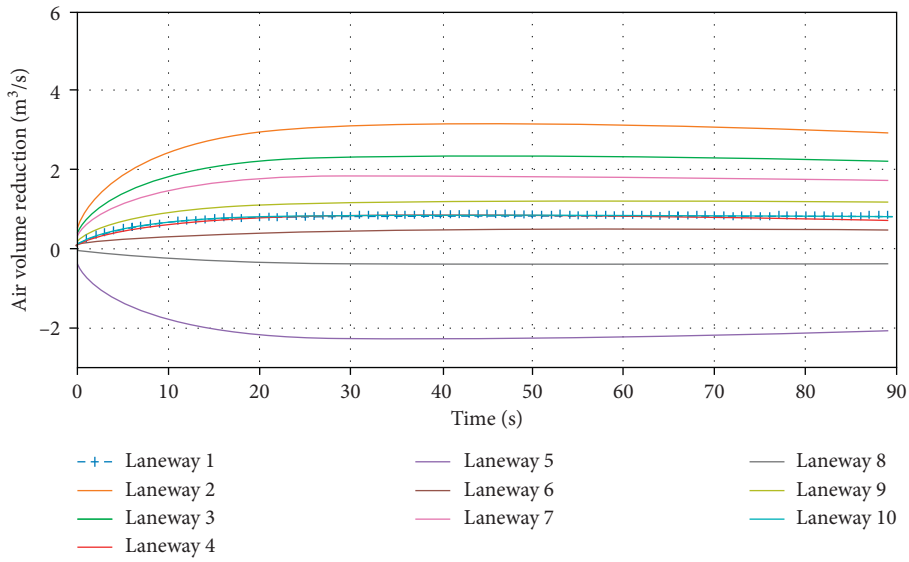


Figure 5: Air volume reduction for each laneway with mine car moving to windward.

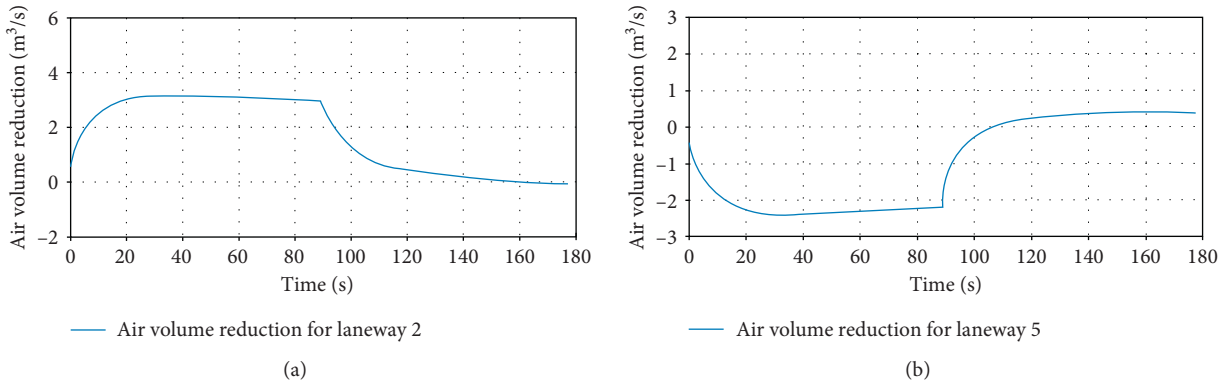


Figure 6: Air volume reduction for laneways 2 (a) and 5 (b) with mine car stops moving to windward.

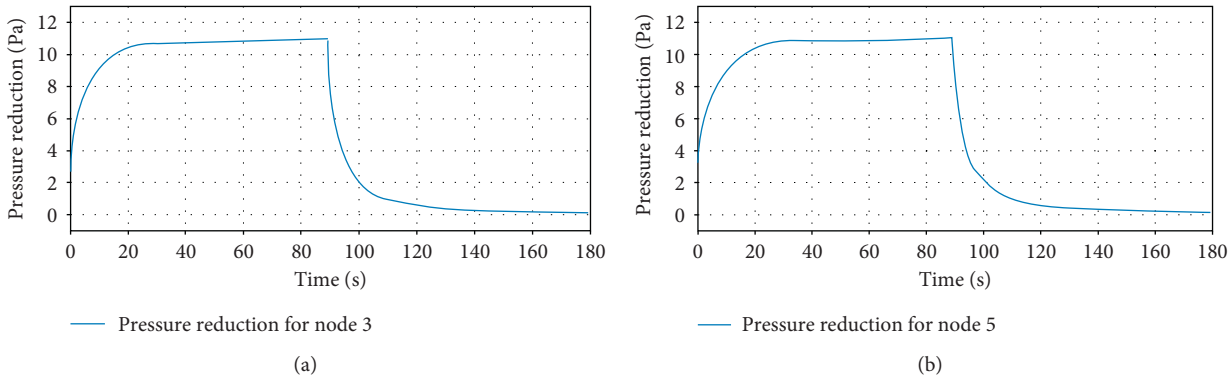


Figure 7: Wind pressure reduction for nodes 3 (a) and 5 (b) with mine car stops moving to windward.

applied to sort the laneway ventilating resistance, based on which 70 residual tree branches and the corresponding minimum spanning tree of the network graph are generated. By adding each of the residual tree branches to the minimum spanning tree, 70 independent circuits are created, respectively. Then, the GIS based nonsteady network model is applied to conduct the ventilation simulation. We obtained the air volume of each laneway branch changing with time. When the ventilation of Xinqiao coal mine reaches the steady state, the result is almost the same with the state based

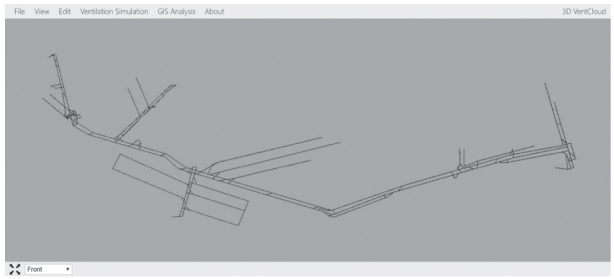


Figure 8: The 3D laneway network model of Xinqiao coal mine.

Table 5: Some of the ventilation attribute data for Xinqiao coal mine laneway.

Laneway ID	Start node ID	End node ID	Wind drag ($N{\cdot}s^2/m^8$)	Section area (m^2)	Section perimeter (m)	Laneway length (m)
1	1	2	122.336	18.14	15.78	32
2	3	4	114.535	18.67	15.46	40.10
3	5	6	145.384	18.28	15.23	3.30
4	7	8	0.00016	24.00	20.37	612
5	9	10	0.00043	14.06	15.89	584

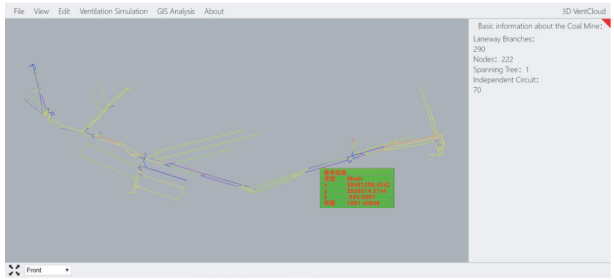


Figure 9: The 3D visualization of mine ventilation simulation result.

on the traditional steady ventilation network model. This study takes a few seconds at the web end to get the simulation result, it consumes 4.72 s in our latest experiment, which is expected to provide fast or real-time decision support for afterwards site applications.

The 3D geometric model of Xinqiao laneway network is rendered based on the simulation results and visualized on 3D VentCloud, as can be seen in Figure 9. The web system supports user interaction such as the geospatial query function. For example, the spatial and attribute features will be shown on web page by clicking on each laneway branch.

5. Conclusions

This paper investigated the unsteady mine ventilation simulation model based on GIS technology and developed a prototype web system to implement the online ventilation simulation, which can better adapt and promote the construction of intelligent mine ventilation.

Specifically, by integrating the GIS technology and unsteady ventilation simulation model, this study provides a unified GIS based ventilation network model, which can effectively simulate the dynamic changes of the unsteady ventilation state as well as the steady ventilation state. Besides, a prototype web system is developed based on the proposed algorithm, which can provide fast simulation result and location based information. The result is demonstrated to be effective in simulating the unified ventilation result, and the simulation system is expected to provide the online and real-time decision making support for coal mine safety production.

However, the ventilation system belongs to a complex and dynamic 3D system. Mine accidents are occasionally happened inside the laneway network. The future work should comprehensively consider the inner structure and 3D characteristics of the laneway network as well as the airflow, and the quantitative accuracy of the model is also worth study in the future.

Data Availability

The data used in this study are confidential. The readers can access the data by sending e-mail to the corresponding author.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was financially supported by the National Key Research and Development Program of China (Grant no. 2016YFC0803108) and the National Natural Science Foundation of China (Grant no. 51774281).

References

[1] C. Ren and S. Cao, "Implementation and visualization of artificial intelligent ventilation control system using fast prediction models and limited monitoring data," *Sustainable Cities and Society*, vol. 52, pp. 101–860, 2020.

[2] C.-H. Chen, F. Song, F.-J. Hwang, and L. Wu, "A probability density function generator based on neural networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 541, Article ID 123344, 2020.

[3] S. Wang, B. Liu, and S. Liu, "Computer simulation of unsteady airflow processes in mine venilation networks," *Journal of Liaoning Technical University (Natural Science)*, vol. 19, no. 5, pp. 449–453, 2000.

[4] G. Wang, H. Wang, H. Ren et al., "2025 Scenarios and development path of intelligent coal mine," *Journal of China Coal Society*, vol. 43, no. 2, pp. 295–305, 2018.

[5] W. Dziurzyński, A. Krach, and T. Pałka, "A reliable method of completing and compensating the results of measurements of flow parameters in a network of headings," *Archives of Mining Sciences*, vol. 60, no. 1, pp. 3–24, 2015.

[6] W. Dziurzyński and E. Krause, "Influence of the field of aerodynamic potentials and surroundings of goaf on methane hazard in longwall N-12 in seam 329/1, 329/1-2 in "krupiński" coal mine," *Archives of Mining Sciences*, vol. 57, no. 4, pp. 819–920, 2012.

[7] C.-H. Chen, "A cell probe-based method for vehicle speed estimation," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E103.A-A, no. 1, pp. 265–267, 2020.

[8] S. Maleki, "Application of VENTSIM 3D and mathematical programming to optimize underground mine ventilation network: a case study," *Journal of Mining & Environment*, vol. 9, 2018.

[9] D. Zielinski, B. Macdonald, and R. Kopper, "Comparative study of input devices for a VR mine simulation," in *Proceedings of the IEEE Virtual Reality (VR)*, Minneapolis, MN, USA, March 2014.

[10] Q. Zhag, Y. Yao, and J. Zhao, "Status of mine ventilation technology in China and prospects for intelligent development," *Coal Science and Technology*, vol. 48, no. 2, pp. 97–103, 2020.

[11] C.-H. Chen, F.-J. Hwang, and H.-Y. Kung, "Travel time prediction system based on data clustering for waste collection vehicles," *IEICE Transactions on Information and Systems*, vol. E102.DD, no. 7, pp. 1374–1383, 2019.

[12] J. Zhang and W. Gao, "Mine ventilation information system based on 3D GIS," *Journal of Liaoning Technical University (Natural Science)*, vol. 31, no. 5, pp. 634–637, 2012.

[13] H. Liu, S. Mao, M. Li, and S. Wang, "A tightly coupled GIS and spatiotemporal modeling for methane emission simulation in the underground coal mine system," *Applied Sciences*, vol. 9, no. 9, p. 1931, 2019.

[14] B.-R. Li, M. Inoue, and S.-B. Shen, "Mine ventilation network optimization based on airflow asymptotic calculation method," *Journal of Mining Science*, vol. 54, no. 1, pp. 99–110, 2018.

[15] P. Wang, B. Lu, and L. I. U. Xu, "Two and three dimensional mine ventilation simulation system based on GIS and AutoCAD," *Safety in Coal Mines*, vol. 47, no. 12, pp. 90–92, 2016.

[16] J. Suh, S.-M. Kim, H. Yi, and Y. Choi, "An overview of GIS-based modeling and assessment of mining-induced hazards: soil, water, and forest," *International Journal of Environmental Research and Public Health*, vol. 14, no. 12, p. 1463, 2017.

[17] S. Choi, M. O. Karslıoğlu, and N. Demirel, "Development of a GIS-based monitoring and management system for underground coal mining safety," *International Journal of Coal Geology*, vol. 80, no. 2, pp. 105–112, 2009.

[18] W. Nyaaba, S. Frimpong, and K. A. El-Nagdy, "Optimisation of mine ventilation networks using the Lagrangian algorithm for equality constraints," *International Journal of Mining, Reclamation and Environment*, vol. 29, no. 3, pp. 201–212, 2015.

[19] X. Wang, X. Liu, Y. Sun et al., "Construction schedule simulation of a diversion tunnel based on the optimized ventilation time," *Journal of Hazardous Materials*, vol. 165, no. 1–3, pp. 933–943, 2009.

[20] Y. Liang, J. Zhang, T. Ren, Z. Wang, and S. Song, "Application of ventilation simulation to spontaneous combustion control in underground coal mine: a case study from Bulianta colliery," *International Journal of Mining Science and Technology*, vol. 28, no. 2, pp. 231–242, 2018.

[21] S. G. Wang, J. R. Wang, and L. Hong, "Mathematical model for solving ventilation networks during conventional ventilation and emergency conditions," *Journal of Liaoning Technical University*, vol. 22, no. 4, pp. 436–438, 2003.

Hindawi
Discrete Dynamics in Nature and Society
Volume 2020, Article ID 6243085, 7 pages
https://doi.org/10.1155/2020/6243085

Review Article

A Virtual Reality Platform for Safety Training in Coal Mines with AI and Cloud Computing

Mei Li,[1] **Zhenming Sun**,[2] **Zhan Jiang**,[1] **Zheng Tan**,[3] **and Jinchuan Chen**[3]

[1]*Institute of Remote Sensing and Geographical Information System, Peking University, Beijing 100871, China*
[2]*School of Energy and Mining Engineering, China University of Mining and Technology (Beijing), Beijing 100083, China*
[3]*Beijing Longruan Technologies, Beijing 100190, China*

Correspondence should be addressed to Zhenming Sun; sun@cumtb.edu.cn

Received 25 June 2020; Revised 20 September 2020; Accepted 26 September 2020; Published 23 October 2020

Academic Editor: Chi-Hua Chen

Coal mining, regarded as a high-risk industry, has a strong demand for virtual reality (VR) to fulfill safety and emergency rescue training. In the past ten years, VR technology has significantly improved miner training on both the hardware and software side. However, it still has some drawbacks, such as expensive and unsuitable hardware, lack of satisfactory user experience, without direct browser access, and lack of humanized and intelligent design. To solve these problems, a cloud-based VR system is designed for the training of coal miners in this paper. The system, with browser/client architecture, includes eight modules demonstrating the full procedure of an underground coal mine. The online cloud-rendered video streaming is adopted to provide enough computing and rendering power and hence a better browser-based user experience. Furthermore, game artificial intelligence (AI) is also introduced into the system to increase the emotional exchange between the system and users. Unlike traditional VR training software, this system designs two virtual miners to enhance the experience of trainees. The first virtual miner is a task-oriented non-player-character (NPC) which conveys general knowledge about the mine and guides the users in visiting the underground work sites. The second virtual miner is a disaster-oriented character which prepares the users for typical disasters. The system has been successfully implemented in a laboratory environment, and its performance has been validated. Yet, further practices are needed to stimulate more innovative applications of VR-based miner training and disaster drilling.

1. Introduction

Coal mines, with dangerous working environments and complex production systems, have a strong demand for virtual reality (VR). VR technology has its unique advantages in several fields related to coal mining, such as miner safety training and emergency rescue drilling, besides drifting-producing process simulation, disaster scenario simulation, and mechanical operation training [1]. The fifth-generation (5G) cellular, artificial intelligence (AI), data analytics, Internet of Things (IoT), and cloud computing will provide the information infrastructure required for the intelligent mine construction in the next three to five years. With these emerging technologies, the next-generation VR systems for underground coal mining will be achieved, which will improve the professional adaptation process and occupational health and safety situation, especially for the young miners [2].

The United States, Australia, and the UK have used VR as a training environment for mining simulation, accident reconstruction and investigation, education, and safety training in the past two decades. The UK, in particular, has a long history of developing and using VR technology in coal mine safety training. The VR products, such as SafeVR and Vroom, are very famous for open-pit truck operator training [3]. The researchers of the National Institute for Occupational Safety and Health (NIOSH) explored how the mining industry could effectively use "serious games" for fire-escape training. Spokane Research Laboratory (SRL) developed fire-escape training software for a mine safety training course. The software allowed four trainees to work together in a virtual environment through a computer network. The

trainees responded favorably and showed a marked improvement in locating the proper evacuation routes on subsequent trials [4]. The University of New South Wales developed iCinema (an advanced visualization and interaction environment), which included a 360° immersive theater environment with 18 modules for improving teaching and learning activities in mining engineering [5]. The University of Queensland Experimental Mine (UQEM) laboratory has been used by students for experiments since 1956. This institution developed several VR applications, including VR-based teaching modules, drill rig, Instron UCS rock testing simulations, and VR-based ventilation simulations [6].

In China, universities, large mining companies, and local governments utilize the VR hardware and software for safety and emergency rescue training [7]. Those VR systems provide the desktop versions of character "Q and A" sessions. The VR hardware is composed of a projection-based panoramic display system, the infrared tracking stereoscopic glasses, the VR headset display, a pad, and the other devices. Furthermore, the 3D visualization system is very popular in coal mines with the functions of the 3D geological model, ventilation simulation, real-time data monitoring, and emergency response.

In recent ten years, many coal mine enterprises established underground industrial networks and tried to apply cloud computing, big data, Internet of Things, and even AI technology into mining, ventilation, safety, and education projects. There has been a shift from traditional mining to intelligent mining. A three-tier IT framework with automatic control system, production execution system, and enterprise management system is established. Based on cloud computing techniques, big data related to underground coal mining can be analyzed quickly, and critical data associated with user requirements can be extracted precisely. With the established platform, the danger-spotting and subsequent decision-making capabilities of users can be greatly enhanced to ensure underground mining safety [8].

Although VR is an important part of intelligent mine construction, there are still some problems with these applications that need to be tackled. Firstly, most of the existing VR-based training systems lack a good learning experience. Although these systems adopt teaching and learning functions via a single-player mode and an ask-answer pattern, there is no complex AI man-machine interaction. VR training systems always focus on professional computation or disaster 3D simulation or text and voice presentation. The learning process needs to be improved. Furthermore, the advantages of immersion, interaction, and imagination of VR technology have not been fully utilized yet [9]. Secondly, traditional VR-based training systems are standalone or client/server versions, without direct browser access. Cloud rendering will become the mainstream to support various light terminals, such as PCs, mobile phones, tablets, and headsets. These technologies will enable real-time transmission and enhance cyber-physical and social experiences in a fully immersive environment [10]. Hence, VR training software should adopt gamification and online learning, which have become a tendency in education.

2. Methodology

To address the abovementioned problems and train the coal miners, a VR-based system using the Unreal Engine is designed and developed in this paper. Contrary to the traditional standalone or client/server VR software, this system adopts browser/client architecture and uses cloud rendering and Game AI technologies. It aims to enhance users' learning experience, strengthen their memory, and improve safety awareness. The system uses the Unreal Engine from Epic Games, which includes a rendering engine (renderer), Blueprint scripting language, animation, artificial intelligence (AI), networking, terrain, streaming, and scene graphs. Figure 1 illustrates the system interface, which has eight modules as follows:

(i) Geological environments, such as procedure of coal formation, geology condition, coal seam, and geology structures

(ii) Geographical environment and surface plants

(iii) Underground roadway arrangement

(iv) Mining area and working face design

(v) Coal mine production system, which includes ventilation, transportation, belt, lifting, power supply, water supply, and drainage

(vi) Mining processes, including coal mine blasting and longwall mining processes

(vii) Coal mine disaster (gas leakages, roof falls, and fire events) simulation

(viii) Underground safety monitoring system (methane, wind speed, temperature, and carbon monoxide), self-rescue, and emergency response

In this paper, we use three technologies, namely, cloud rendering architecture, 3D modeling, and AI behavior design to achieve multiuser online operation and improve the interaction and immersion of VR devices.

The browser/client architecture of cloud rendering means that each user connects to the server and accesses 3D streaming over the cloud. The cloud servers handle all events and the state of the virtual world and transfer the VR render streaming to the client. These servers can be virtual servers or server clusters. They take full advantage of virtually infinite computing power to create, capture, and compress high-resolution rendered streaming, which is sent to the client with low performance and limited computing power [11].

Compared with other popular Web3D technologies, such as HTML5 and WebGL, cloud rendering has incomparable advantages mainly because it has no special or expensive requirements on the client side. It allows users to choose different platforms, including PCs, laptops, tablets, and smartphones. Furthermore, since Wi-Fi 6 and 5G will provide sufficient network bandwidth, online cloud rendering will deliver adequate computing and rendering power. With the advancements in the Internet infrastructure, the users can interact with VR programs through the network on the client with high-quality videos and low-response delays [12].

Figure 1: The interface of the virtual reality system for underground coal mine.

In this B/S architecture of the cloud rendering system, the system is divided into three parts: portal server, render server, and client, as shown in Figure 2.

The portal server is responsible for user login, finding the appropriate server according to the user's choice, and controlling the server to start the virtual machine, i.e., render server. The client and the render server establish a connection when the portal server notifies the client with the address of the virtual machine. The user can run and interact with the 3D program in the remote server through the network.

The render server is responsible for processing the input events of the client sourced by mouse or keyboard, as well as translating and sending them to the interaction module of the 3D program for further parsing. The 3D scene images are captured, encoded, and sent to the client as a video streaming. The render server program includes three modules, namely, basic function, Game AI, and advanced function. The VR basic function module provides 3D scene management, sound, particle effect, physical simulation, and other basic functions. The Game AI module uses a scripting language to design a behavior tree for virtual miners. The VR advanced module builds complex computation models and simulates disasters, such as fire numerical simulation and escaping route computation.

The client is responsible for displaying the received streaming of the 3D scene and sending user inputs, such as keyboard and mouse events, to the server.

We use 3ds Max to build 3D scenes of the underground and the surface. Then, we import the 3D models into the Unreal Engine editor. We adopt a series of material tools to apply 3D models, such as a normal map, diffusion map, metal map, and texture map. These materials are like the "paint" applied to the model. Then, we set the lights to illuminate our scenes. From the computer-aided design (CAD) or geographic information system (GIS) files, we automatically build 3D geological and underground laneway models. We use the Blueprint scripting language to program the characters' animation, the AI behavior tree, and the user interface (UI).

AI in games or "Game AI" is a module of a 3D engine. Traditional AI typically demonstrates some of the following behaviors associated with human intelligence: planning, learning, reasoning, problem-solving, knowledge representation, perception, motion, and manipulation. The emerging Game AI is becoming a narrow branch differentiated from the traditional AI. Game AI refers to the virtual man, also called non-player characters (NPCs), controlled by the computer, which simulates the intelligent behaviors of human beings or other creatures, providing users with a reasonable challenge to overcome [13].

Figure 3 shows the behaviors of the two NPCs we designed. These NPCs can simulate the behaviors of miners and guide young miners and/or students to understand the whole mining production, experience the underground scene, and improve the relationship between users and computer programs during the training, conversation, or other tasks. Besides, these AI characters are designed to automatically find the local navigation grid path.

The first NPC is a task-oriented character. When the user meets it somewhere on the ground, he will provide general knowledge about the coal mines and guide the user to descend a well from a cage and visit the underground work sites, such as longwalls or heading faces. Once the user completes the corresponding task and provides feedback, the NPC releases the next task. Those tasks are performed before entering the underground mine, preparing the miners to look for the entrance, enter the cage, and visit the underground working face. The second NPC is a disaster-oriented character which guides the user to experience typical disasters, such as gas explosions, roof falls, and fire accidents.

To realize the NPC's behavior, the blueprint is used to build a behavior tree graph, which adds and connects a series of nodes. A behavior tree, mainly composed of a leaf node

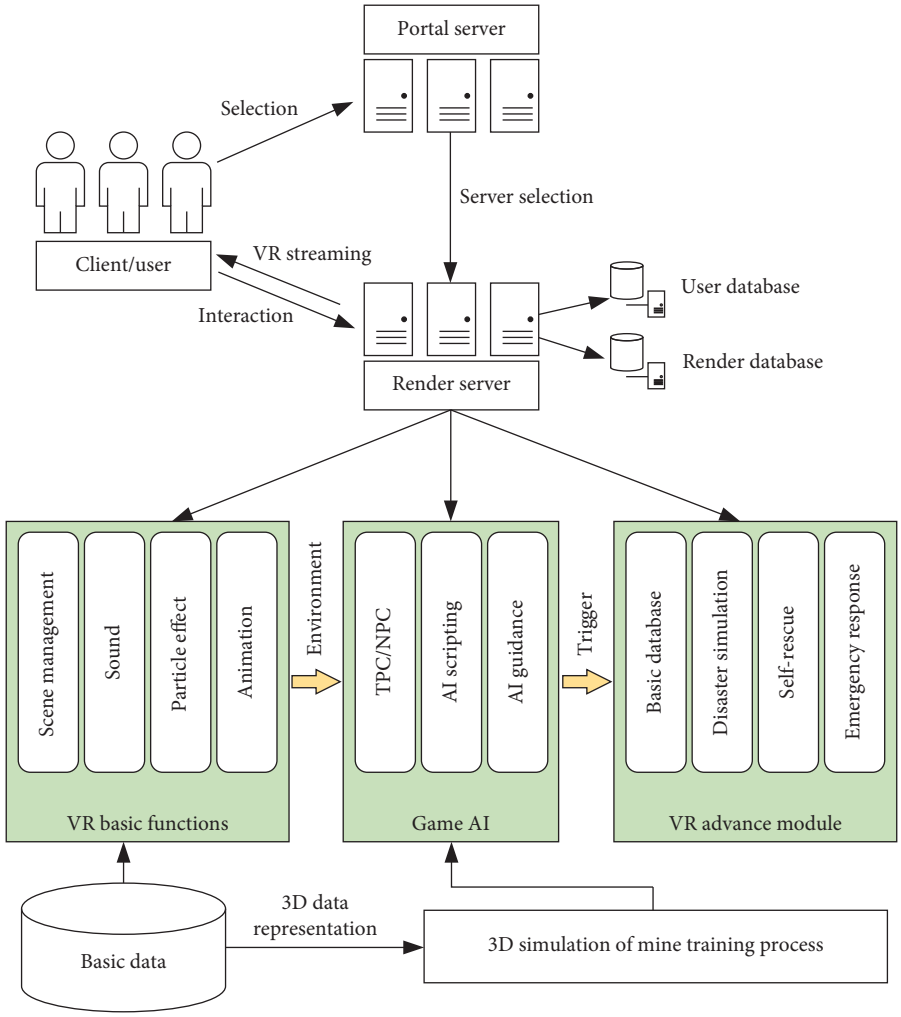


Figure 2: The architecture of the cloud rendering system.

and composite node, is essentially an acyclic graph [14, 15]. We use composite tasks and design nodes of selectors, sequences, parallels, and decorators. Figure 4 shows an example of the second NPC's behavior tree, where selector, sequence, and parallel are used to trigger accidents. A selector is a branch task that runs the behavior of each of its children. It will immediately return a success status code when one of its children runs successfully. Once one of the child nodes is successful, it returns to the selector, and the rest of the nodes are not executed. A sequence executes the node task sequentially instead. A sequence node tries to execute each child in turn until all children have reported that they have executed successfully. Once a task fails, the sequence will return to the upper node. In this case, the remaining tasks are not executed. A parallel is a composite task that handles concurrent behaviors. It is a special branch task that runs all children during execution time. The behavior tree allows parallel node tasks, which might be state machines. Multiple state machines can be executed in parallel. For example, when the second NPC's behavior tree is triggered, it first judges whether it is dangerous or not and then runs the user and sends alerts. Following these, it gives the on-site emergency solution by a series of animation actions. The sequence executes the disaster animation in turn and tests whether the user's reaction is correct.

3. Results and Discussion

3.1. System Implementation. The cloud rendering cluster is composed of two virtual rendering servers with a configuration of 4-core 8G M60 CPU, 32 GB RAM, and 40 Mbps BGP bandwidth. The experiments have shown that the hardware can support up to seven players at the same time. A client PC should have at least Intel i7 CPU and 16 GB RAM. The system has been tested by undergraduate students in the VR Teaching and Experiment Laboratory, China University of Mining & Technology, Beijing. The laboratory also has a 120-degree 2-channel VR projector and an infrared laser tracking system as shown in Figures 5 and 6, respectively. The system can run either on the immersive VR hardware or the PC headset.

3.2. AI Guidance. The virtual roaming function of the NPC will help the user to walk around buildings, enter the mine through a cage, and visit the longwall mining face or the

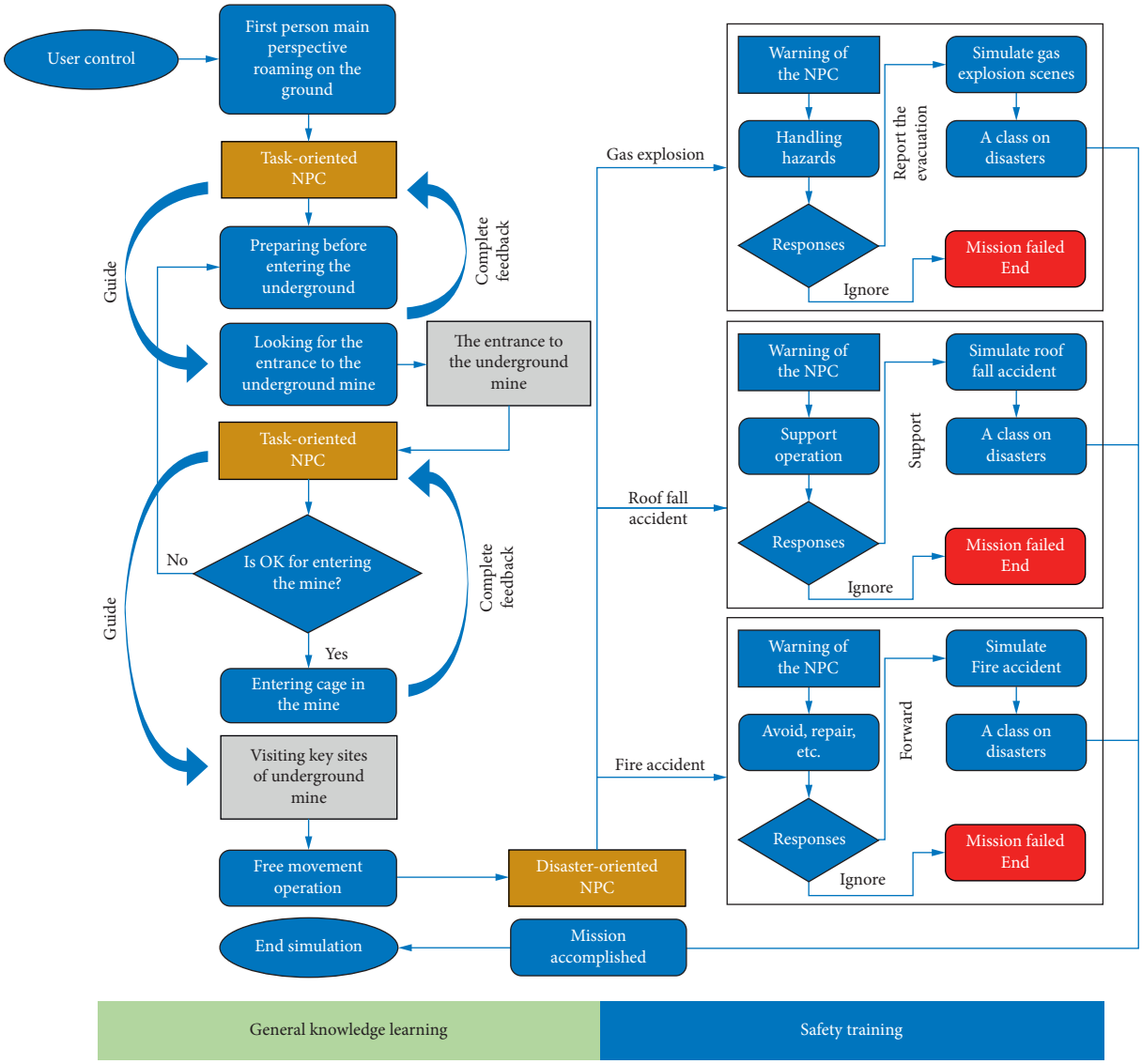


Figure 3: VR mission design of Game AI NPCs.

blasting place. It will also introduce the ventilation system, transportation system, water supply, drainage system, and power supply system. This function enables the new miners or students to have an overall understanding of the mine.

When a user is wandering in the roadway, the accident guidance function of the NPC will start a dialogue with them to detect their behavior and provide information on illegal operations. The system will then trigger disaster animation depending on the location, and the NPC will demonstrate the escape routes for self-rescue.

3.3. Disaster Simulation. In our system, mining accidents are limited to gas explosions, roof falls, and fire events. The disaster simulation module integrates various 3D animation, audio, particle effect, and physical-mechanical models of the 3D engine. The 3D model is to establish the models of mines, terrain, tunnels, and NPCs. The audio function is to create sounds as realistic as possible considering the dimension of the source, such as footsteps when walking on the grass and the roadway, wheezing when running, and the mechanical sound of an opening cage. By analyzing the disaster/accident reasons, features, and occurrence conditions, a disaster simulation knowledge library is constructed. This library contains a brief description of prevention and control measures, emergency plans, site disposal programs, mitigation, and postdisaster evaluation. It is combined with system guidance to show the potential consequences of disasters.

3.3.1. Gas and Coal Dust Explosion. This simulation is to show the boundary conditions of gas and coal dust explosion as illustrated in Figure 7. It helps users to perceive the disaster from the vision, auditory, and other aspects and understand the movement and destruction of the wind that is mixed with dust. Besides, some emergency measures, such as collaborative gas extraction, are fully displayed.

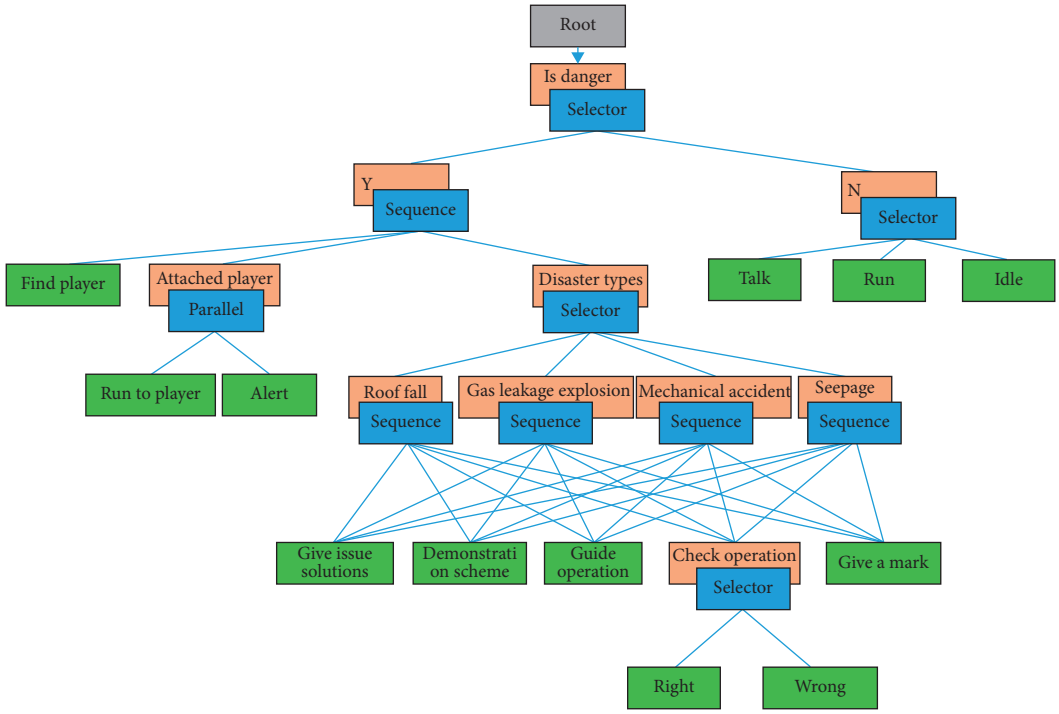


Figure 4: The behavior tree of the disaster-oriented NPC.

Figure 5: The professor introducing the system to the visitors on the 120-degree 2-channel VR projector.

Figure 6: The overview of VR Teaching and Experiment Laboratory.

3.3.2. Roof Fall and Wall Collapse. Through this simulation, the user will understand the main factors of roof fall and wall collapse as shown in Figure 8. It can simulate various rock burst and ground pressure disasters, introducing the form of disasters, their causes, and the potential methods of disaster prevention and control.

3.3.3. Fire Event. The user will see the main forms of fires, such as the external fires that are mostly caused by open flames, blasting, and short circuits, as well as the internal ones caused by the combustion of coal or other flammable substances due to their oxidation or burn. The VR system can simulate the conditions of various fires, illustrate their severity, and therefore inform decision making for disaster prevention. The system also shows the ways of treatment, including the closed grouting and nitrogen injection methods.

Figure 7: Gas explosion animation.

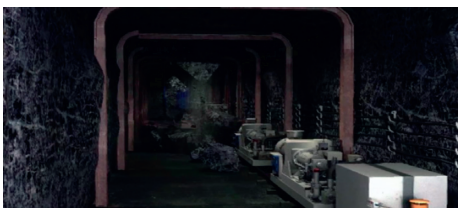


Figure 8: Roof fall animation.

4. Discussion

The VR system uses cloud rendering and AI behavior trees to enhance training performance. It can realistically simulate

the complex underground coal mine environment and provide the function of login, roaming, virtual simulation, and operation evaluation. However, due to intellectual property rights, only a small part of the functions is made accessible to the public through the website. To date (August 1, 2020), the system has been visited 20, 853 times and the training system drew perfect user reviews (5.0/5.0) on the National Virtual Simulation Experiment Teaching Project Sharing Platform [16]. According to the experiment score data, 1355 people have finished the experiment and the excellent score rate was 46.0%. The students commented that the system could give them a more intuitive understanding of the underground environment and disasters without visiting the real underground coal mines. Despite satisfactory performance and remarkable feedback, the current cloud render hardware needs to be upgraded to a larger cloud rendering cluster to provide a quicker response and hence a better experience. Furthermore, a performance test is needed to estimate the amount of concurrency and runtime efficiency. The system functions also need more research, where a more detailed behavior tree design can improve the user experience. Combined with the numerical computation, the system can be turned into a testbed to support the ventilation design, emergency drilling, and other systems and fields. We expect more innovative applications of VR-based miner training and disaster drilling in the near future.

5. Conclusions

VR, as a part of intelligence mine, provides the new tool for work procedure management, occupational health and disaster simulation, and visualization. The VR system provides macroscopic and microscopic underground 3D scenes for coal mine and displays information such as personnel positioning, safety monitoring, and equipment management. Combined with cloud rendering, it can achieve more realistic and more detailed 3D visualization effects. Considering the advantages of cloud rendering and AI, this paper designs a cloud-based VR system for mine safety training. AI guidance is adopted to roam underground and simulate accidents, such as gas explosions, roof falls, and fire events. We believe that the results of this study will provide a useful tool for coal mine safety training, avoiding the potential hazards of the extreme underground environment.

Data Availability

No data were used to support the findings of this study. The system will be considered to be openly accessed afterward.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was financially supported by the National Key Research and Development Program of China (grant no. 2016YFC0803108). It was also supported by the Key Research and Development Program of Inner Mongolia titled "Technology and application on digital mine resource management and ecological environment monitoring, 2015–2019."

References

[1] J. Tichon and R. Burgess-Limerick, "A review of virtual reality as a medium for safety related training in the minerals industry," *Journal of Health & Safety Research & Practice*, vol. 1, no. 3, pp. 33–40, 2011.

[2] A. Grabowski and J. Jankowski, "Virtual reality-based pilot training for underground coal miners," *Safety Science*, vol. 72, no. 72, pp. 310–314, 2015.

[3] D. Schofield, B. Denby, and R. Hollands, "Mine safety in the twenty-first century: the application of computer graphics and virtual reality," in *Mine Health and Safety Management*, M. Karmis, Ed., pp. 153–174, Society of Mining, Metallurgy, and Exploration, Inc., Englewood, CO, USA, 2001.

[4] T. J. Orr, L. G. Mallet, and K. A. Margolis, "Enhanced fire escape training for mine workers using virtual reality simulation," *Mining Engineering*, vol. 61, no. 11, pp. 41–44, 2009.

[5] T. Barker, "Images and eventfulness: expanded cinema and experimental research at the University of New South Wales," *Studies in Australasian Cinema*, vol. 6, no. 2, pp. 111–123, 2012.

[6] M. S. Kizil, A. P. Kerridge, and M. G. Hancock, "Use of virtual reality in mining education and training," in *Proceedings of the 2004 CRC Mining Research and Effective Technology Transfer Conference*, pp. 1–7, Noosa Heads, Australia, June 2004.

[7] M. Li, Z. M. Sun, P. Y. Lyu, J. Chen, and S. Mao, "Study on key technology of multiplayer virtual reality training platform for fully-mechanized coal mining face," *Coal Science and Technology*, vol. 46, no. 1, pp. 156–161, 2018.

[8] Y. Wu, M. Chen, K. Wang, and G. Fu, "A dynamic information platform for underground coal mine safety based on internet of things," *Safety Science*, vol. 113, pp. 9–18, 2019.

[9] R. Mitra and S. Saydam, "Can artificial intelligence and fuzzy logic be integrated into virtual reality applications in mining?" *Journal of the Southern African Institute of Mining and Metallurgy*, vol. 114, no. 12, pp. 1009–1016, 2014.

[10] E. Bastug, M. Bennis, M. Medard, and M. Debbah, "Toward interconnected virtual reality: opportunities, challenges, and enablers," *IEEE Communications Magazine*, vol. 55, no. 6, pp. 110–117, 2017.

[11] R. Shea, J. Jiangchuan Liu, E. C.-H. Ngai, and Y. Cui, "Cloud gaming: architecture and performance," *IEEE Network*, vol. 27, no. 4, pp. 16–21, 2013.

[12] C. Y. Huang, C. H. Hsu, Y. C. Chang, and K.-T. Chen, "GamingAnywhere: an open cloud gaming system," in *Proceedings of the 4th ACM Multimedia Systems Conference*, pp. 36–47, New York, NY, USA, February 2013.

[13] G. N. Yannakakis and J. Togelius, "A panorama of artificial and computational intelligence in games," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 7, no. 4, pp. 317–335, 2015.

[14] I. Millington and J. Funge, *Artificial Intelligence for Games*, CRC Press, Boca Raton, FL, USA, 2009.

[15] W. J. He, *Design and Implementation of Game AI Based on Behavior Tree*, Chengdu University of Technology, Chengdu, China, 2018.

[16] National Virtual Simulation Experiment Teaching Project Sharing Platform, September 2020, http://www.ilab-x.com/details/v5?id=5022&isView=true.

Hindawi
Discrete Dynamics in Nature and Society
Volume 2020, Article ID 8782450, 12 pages
https://doi.org/10.1155/2020/8782450

Research Article

Coal Mine Gas Safety Evaluation Based on Adaptive Weighted Least Squares Support Vector Machine and Improved Dempster–Shafer Evidence Theory

Zhenming Sun [1] **and Dong Li** [2]

[1]*School of Energy and Mining Engineering, China University of Mining and Technology (Beijing), Beijing 100083, China*
[2]*Beijing Longruan Technologies, Beijing 100190, China*

Correspondence should be addressed to Zhenming Sun; sun@cumtb.edu.cn

Received 25 June 2020; Revised 3 September 2020; Accepted 26 September 2020; Published 20 October 2020

Academic Editor: Chi-Hua Chen

Gas safety evaluation has always been vital for coal mine safety management. To enhance the accuracy of coal mine gas safety evaluation results, a new gas safety evaluation model is proposed based on the adaptive weighted least squares support vector machine (AWLS-SVM) and improved Dempster–Shafer (D-S) evidence theory. The AWLS-SVM is used to calculate the sensor value at the evaluation time, and the D-S evidence theory is used to evaluate the safety status. First, the sensor data of gas concentration, wind speed, dust, and temperature were obtained from the coal mine safety monitoring system, and the prediction results of sensor data are obtained using the AWLS-SVM; hence, the prediction results would be the input of the evaluation model. Second, because the basic probability assignment (BPA) function is the basis of D-S evidence theory calculation, the BPA function of each sensor is determined using the posterior probability modeling method, and the similarity is introduced for optimization. Then, regarding the problem of fusion failure in D-S evidence theory when fusing high-conflict evidence, using the idea of assigning weights, the importance of each evidence is allocated to weaken the effect of conflicting evidence on the evaluation results. To prevent the loss of the effective information of the original evidence followed by modifying the evidence source, a conflict allocation coefficient is introduced based on fusion rules. Ultimately, taking Qing Gang Ping coal mine located in Shaanxi province as the study area, a gas safety evaluation example analysis is performed for the assessment model developed in this paper. The results indicate that the similarity measures can effectively eliminate high-conflict evidence sources. Moreover, the accuracy of D-S evidence theory based on enhanced fusion rules is improved compared to the D-S evidence theory in terms of the modified evidence sources and the original D-S evidence theory. Since more sensors are fused, the evaluation results have higher accuracy. Furthermore, the multisensor data evaluation results are enhanced compared to the single sensor evaluation outcomes.

1. Introduction

China is a country with a large coal consumption and production where a large proportion of the production mines is related to the high gas mines. The gas accident is one of the major problems; hence, it is necessary to investigate and solve this problem for China's coal industry. Coal mine gas safety evaluation has always been a key tool for coal mine safety management. In China, the coal mines are ordered to monitor the gas concentration, carbon monoxide concentration, carbon dioxide concentration, oxygen concentration, dust, wind speed, humidity, temperature, power state, and others by the National Coal Mine Safety Administration [1]. Through monitoring those data automatically and identifying the gas safety state timely in the coal mine, outburst, gas accumulation, and explosion can be effectively prevented. The work has important theoretical significance and practical value for suppressing the gas disasters occurrence [2, 3] and endorsing the safe and sustainable development of the coal industry.

Safety evaluation and risk assessment are important and systematic processes to assess the impact, occurrence, and

consequences of human activities on a system with hazardous characteristics, and they are necessary tools for the company's safety policy. The risk types and data sources are many and various, so are the safety evaluation techniques to assess risks. Therefore, the choice of methods has become more and more important. Presently, safety evaluation techniques can be classified into qualitative and quantitative safety evaluation methods [4, 5].

1.1. Safety Evaluation Techniques

1.1.1. Qualitative Methods. The qualitative safety assessment methods are mainly to carry out qualitative analysis of the production system's process, equipment, facilities, environment, personnel, and management based on experience and intuitive judgment ability. The results of qualitative safety assessment methods are some qualitative indicators, such as the type of an accident and the factors that may lead to the accident. The commonly used qualitative analysis methods [4] include checklist analysis, plant level safety analysis, process risk management audit, failure mode effect analysis, hazard, and operability. The qualitative evaluation process is simple and easy to understand and manage; however, the differences in the professional background and operational capabilities of various participants may lead to differences in safety assessment. For example, the structure of checklist analysis relies exclusively on the knowledge built into the checklists to identify potential problems [6]. If the checklist does not address a key issue, the analysis is likely to overlook potentially important weaknesses.

1.1.2. Quantitative Methods. The quantitative safety assessment methods are to quantify the status of the production system's processes, equipment, environment, facilities, personnel, and management, based on statistical analysis of a large number of experimental results or/and accident data, using obtained indicators or laws (mathematical models). The commonly used quantitative analysis methods [5, 7] include fault tree analysis, event tree analysis, shortcut risk assessment, and maintenance analysis. The quantitative methods can evaluate the system more accurately than the qualitative methods, but they are still not perfect. Take the commonly used fault tree analysis as an example, this is a deductive technique that uses a fault tree to determine the cause of the accident event. All possible accident events are needed to construct the complete fault tree, but it is difficult to assess all possible accident events and their possibilities and consequences.

1.2. Intelligent Methods for Coal Mine Safety Evaluation. With the enormous development of artificial intelligence (AI), more and more practical applications are available with the artificial intelligent algorithm in the field of engineering [8–12], and numerous attempts have also been carried out on coal spontaneous combustion [13, 14], gas explosions [15–18], etc. Moreover, there are some intelligent methods for coal mine safety evaluation to assess risks quantitatively and solve the problem above.

1.2.1. Improved Fuzzy Theory Methods. Sun [19] developed a comprehensive assessment model of coal mine safety risk in terms of the Fuzzy TOPSIS and integration operator technique. Dai [20] presented a method to use the gas density data by leveraging the fuzzy synthetic evaluation model, and an algorithm to select the weights assignment proposals. Peng [21] introduced linguistic intuitionistic fuzzy numbers to depict the necessary evaluation information. Wang [22] estimated and ranked all of these risk factors through the fuzzy analytic hierarchy process including managerial, environmental, individual, and operational criteria to develop a management model and direct the safety managers in mining procedure.

1.2.2. Improved Swarm Intelligence. He [23] integrated an ant colony algorithm with neural networks to develop a neural network security assessment model utilizing an ant colony algorithm to train the neural network weights. Li [24] optimized the neural net model of the right value (threshold) to overcome the neural net easily falling into the local minimum through quantum genetic algorithm.

1.2.3. Other Techniques. An improved factorization-machine supported neural networks (FNN) structure was designed by Zhang [25]. The fuzzy neurons of the improved FNN have decision-making and control properties with further enhanced error correction performance making the entire system adaptable and stable. Wang [26] integrating the gray correlation technique and the new gray correlation degree method introduced a dynamic resolution coefficient to decrease the error of the gray correlation technique. Li [27] extracted the causal chain of accidents through Bayesian Network analysis to develop the multilevel forecasting indicator system for safety situations and constructed the multilevel prediction model for the coal mine risk trend by combining rough set theory, Bayesian network, and support vector machine.

The occurrence reasons of gas accidents mostly include the unfavorable monitoring of environmental factors, the insufficient accuracy, and the lack of evaluation systems. Based on the Dempster–Shafer (D-S) evidence theory, a coal mine gas safety evaluation model is proposed to automatically get more accurate safety state information. Various sensor monitoring data were collected from the working face monitoring system and processed by adaptive weighted least squares support vector machine (AWLS-SVM) to obtain the prediction data as the input of the safety assessment model. Then, the gas safety state was divided into some different safety levels, and multisensor data fusion was carried out. By the comprehensive analysis of fusion results, the gas safety assessment would be realized assisting coal mine safety management.

2. Gas Safety Evaluation Model

The data source of the gas safety evaluation model is the monitoring system of a coal mine, including monitoring data of gas concentration, wind speed, dust, temperature, etc. The Pearson correlation was used to find the reasonable correlative sensors.

First, in order to evaluate the safety state of a coal mine after a certain time from now on, the predicted sensor values are acquired using the time series prediction model, and AWLS-SVM has been used as the prediction model in this paper. The predicted sensor values would be the evidence sources used in the next step.

Second, the sensor data need to be integrated using improved D-S evidence theory to produce more consistent, accurate, and useful information for safety assessment. The basic probability assignment (BPA) function, which is the basis of D-S evidence theory, is obtained through the posterior probability modeling technique and the similarity degree is presented for modifying the evidence source to reduce the conflicts and improve the accuracy.

Ultimately, multisensor data fusion is performed based on the introduced fusion rules, which are enhanced by the conflict assignment coefficients to prevent the distortion of evidence sources. The reasonable modification of fusion rules can also enhance the accuracy of fusion results.

The coal mine gas safety evaluation model based on AWLS-SVM and D-S evidence theory is shown in Figure 1.

3. Adaptive Weighted Least Squares Support Vector Machine

The evaluation model developed in this paper is aimed at the coal mine gas safety assessment after a definite time. It is essential to obtain the predicted values of the monitoring variables as the input of the evaluation model. Therefore, an adaptive weighted least squares support vector machine (AWLS-SVM) is proposed based on the weighted least squares support vector machine (WLS-SVM) with the adaptive weights calculated through the distribution characteristics of the discrete points.

3.1. Weighted Least Squares Support Vector Machine. Suykens [28] proposed a WLS-SVM in terms of the least squares support vector machine (LS-SVM). The Lagrange function of its optimization problem can be explained as

$$L(w,b,\xi,\alpha) = \frac{1}{2}w^T w + \frac{1}{2}C\sum_{i=1}^{N} v_i \xi_i^2 - \sum_{i=1}^{N} \alpha_i \left[w^T \varphi(x_i) + b + \xi_i - y_i\right]. \tag{1}$$

In the previous equation, w represents the weight coefficient vector; $\varphi(x_i)$ shows the mapping input to the high-dimensional space; C denotes the regularization parameter; b represents the threshold; x_i is the Lagrange multiplier. Regarding the KKT (Karush–Khun–Tucker) condition, the function eliminates w, ξ_i, and obtains

$$\begin{bmatrix} 0 & l_{1\times N} \\ l_{N\times 1} & R + \frac{1}{C}V \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix}. \tag{2}$$

In the previous equation, $V = \mathrm{diag}(v_1^{-1}, v_2^{-1}, \ldots, v_N^{-1})$ represents the diagonal matrix, $l_{1\times N}$ shows the unit column vector, $R = \{K(x_i, x_j) | i = 1, 2, \ldots, N\}$ denotes the radial basis kernel function matrix, and $y = [y_1, y_2, \ldots, y_N]^T$. Equation (2) can be obtained α and b, inputting test samples to obtain WLS-SVM model as

$$y = \sum_{i=0}^{n} \alpha K(x_i, x) + b. \tag{3}$$

The weight calculation formula is

$$v_i = \begin{cases} 1, & \left|\frac{\xi_i}{\hat{s}}\right| \le s_1, \\ \frac{s_2 - \left|\xi_i/\hat{s}\right|}{s_2 - s_1}, & s_1 < \left|\frac{\xi_i}{\hat{s}}\right| \le s_2, \\ 10^{-4}, & \text{otherwise.} \end{cases} \tag{4}$$

In the previous equation, the values of s_1 and s_2 are 2.5 and 3.0, respectively [28]; $\hat{s}$ represents the standard estimated deviation of the error sequence; and its calculation function is

$$\hat{s} = \frac{\mathrm{IQR}}{2 \times 0.6745}. \tag{5}$$

In the previous equation, IQR represents the difference between the first and third quartiles in the sequence of errors ξ_i from small to large.

3.2. Adaptive Weighted Least Squares Support Vector Machine. In the WLS-SVM algorithm, the weight is mainly used to eliminate the influence of gross error data in the sample, and whether its value is appropriate directly determines the performance of the model. The weights determined by equation (4) are linearly distributed, and the calculation results will include errors. Therefore, this paper adaptively determines the weight of each sample through iterative operations. The weights are adaptively calculated utilizing the distribution characteristics of discrete points divided into two categories: high leverage points far from the input data center, and high residual error points differing greatly from the actual value. The weighting technique in this paper is the key of the AWLS-SVM, the gross errors are judged simultaneously through sample leverage points and residual points, and it can minimize the adverse impacts of the discrete points.

The residual error weight v_i^{ξ} of the i-th sample data is determined as

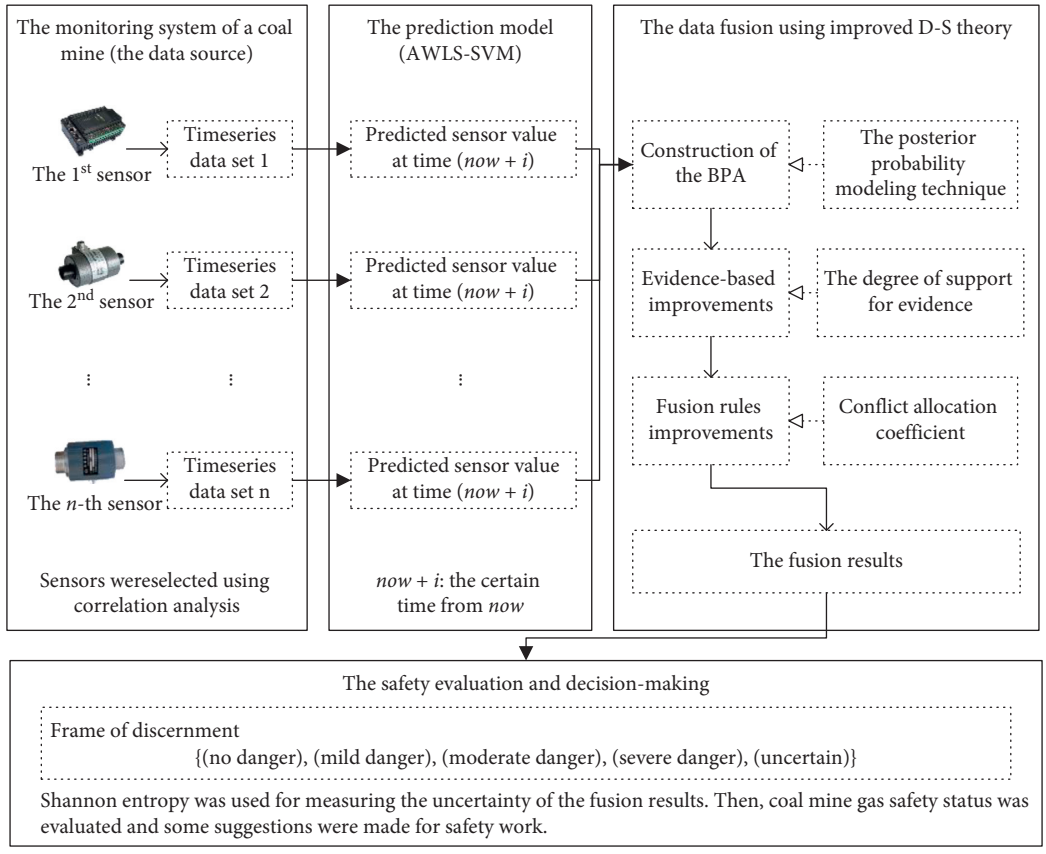


Figure 1: The gas safety evaluation model.

$$v_i^{\xi} = \frac{2}{1 + e^{\xi_i / T}}, \quad i = 1, 2, \ldots, n. \tag{6}$$

where T represents the robust scale estimate of the residual error defined as

$$T = \text{median}\left|\xi_i - \text{median}\left(\xi_i\right)\right|, \quad i = 1, 2, \ldots, n. \tag{7}$$

The leverage weight v_i^x of the i-th sample data is determined as

$$v_i^x = f\left(\frac{\left\|x_i - \text{median}\left(X\right)\right\|}{\text{median}\left(\left\|x_i - \text{median}\left(X\right)\right\|\right)}, c\right), \quad i = 1, 2, \ldots, n,$$

$$f\left(z, c\right) = \frac{1}{\left(1 + |z/c|\right)^2} \tag{8}$$

where $\|\cdot\|$ represents the Euclidean distance, $\text{median}\left(X\right)$ shows the median value of X, x_i denotes the i-th sample data, X represents the vector of all input specimens, and c denotes a constant usually taken as 4 [28, 29].

Comprehensively considering the leverage and residual error weights, the weight v_i of the i-th sample data is determined as

$$v_i = \sqrt{v_i^{\xi} v_i^x}. \tag{9}$$

The algorithm steps of AWLS-SVM are given in Algorithm 1.

3.3. Prediction Model of Monitoring Variables. The steps to build a prediction model of monitoring variables in terms of AWLS-SVM include the following:

Step 1. Collecting the monitoring data of the mine working face, preprocessing the data, and obtaining the learning samples of the model.

For missing data in the time series, the interpolation method is utilized to supplement the missing data. For abnormal data, which are values of zero or beyond the theoretical range, the discarding method is used to delete the abnormal data from the original data set, and the interpolation method is used again as a supplement. The ultimate objective is to prevent data problems resulting in the deviation of the counterintuitive results.

Step 2. Dividing the learning sample into a test set and a training set and selecting the proper fitness function, like the mean square error (MSE), neighborhood average method, and weighted arithmetic mean. The MSE is used in this paper.

Step 3. Using 3-fold cross-validation, performing the regression analysis by WLS-SVM with the training sample data, and determining the fitting residual error ξ of each sample. The initialization weight value v is

Step 1. Based on the modeled sample data, determine the fitting residual of each sample using least squares support vector machine regression
Step 2. Initialize the weight v_i using equations (6), (8), and (9)
Step 3. Perform weighted least squares support vector machine regression on the sample data to obtain a regression model
Step 4. According to the regression model, calculate the residual error ξ of each sample data and recalculate the weight v_i using equations (6), (8), and (9)

Algorithm 1: Algorithm steps.

calculated based on equations (4) and (5) and taking v into equations (2) and (3) to reach the WLS-SVM model.

Step 4. Recalculating the residual error ξ of each sample data based on the regression model. The weight value v is recalculated using equations (6), (8), and (9).

Step 5. Taking the weight value v into equations (2) and (3) to develop the AWLS-SVM as the prediction model.

4. Improved Dempster–Shafer Evidence Theory

Dempster–Shafer (D-S) theory has strong applicability in data fusion; however, there are still some deficiencies in the actual fusion process in dealing with uncertain problems. The high conflicts of uncertain information may make the data fusion results inconsistent with the facts [30], resulting in the inability to assess the event. The problems are mainly manifested in the following three aspects:

One-vote veto problem: when there is a complete contradiction between the pieces of evidence, there will be a veto problem

General conflict problem: when the belief functions of the evidence are very different, unreasonable results appear after fusion

Robustness problem: when the belief functions of the evidence change, the results after data fusion will change drastically

In this study, the enhancement of D-S evidence theory is mostly considered to solve the problem of conflicting evidence sources.

4.1. Basic Principles of D-S Evidence Theory. For reasoning, the uncertain problems, D-S evidence theory has robust adaptability with a simpler reasoning process. The distribution of belief functions and the fusion of evidence are the basic knowledge of D-S evidence theory. The uncertainty of events can be expressed through the frame of discernment and basic probability assignment function.

4.1.1. Frame of Discernment. A set X of possible situations of the event is represented by the frame of discernment with the elements representing the degree of evaluation of the event state. In the gas safety evaluation system, every possible state is known as a hypothesis, and all possible categories constitute a frame of discernment. Hence, the frame of discernment includes all possible results of a particular problem. The frame of discernment can be expressed in

$$X = \{X_1, X_2, X_3, \ldots, \Theta\}, \tag{10}$$

where X_i represents a possible result of the event and Θ denotes the uncertainty.

4.1.2. Basic Probability Assignment (BPA) Function. Suppose that X is a frame of discernment; 2^X represents a power set on X, if m: $2^X \longrightarrow [0, 1]$ and satisfies

$$\sum_{A \in 2^X} m(A) = 1, \quad m(\Theta) = 0. \tag{11}$$

In the previous equation, m is known as the BPA of the discernment frame X and it is also known as the mass function and A represents the element in the discernment frame. For $\forall A \subseteq X$, $m(A)$ shows the basic belief indicating the level of trust in proposition A.

4.1.3. Belief Function. If there are $A \in P(X)$ and $B \in A$, the function Bel is defined as

$$\mathrm{Bel}(A) = \sum_{B \in A} m(B). \tag{12}$$

In the previous equation, Bel shows the belief function, and equation (11) is the sum of the possibilities of all the subsets of A representing the overall degree of trust in A; hence, it can be inferred that $\mathrm{Bel}(\Theta) = 0$ and $\mathrm{Bel}(X) = 1$. The belief function shows the trust degree of a certain thing. It is incomplete and untrustworthy to only use the belief function to explain the possibility of an event.

4.1.4. Likelihood Function. In D-S evidence theory, the likelihood function is a measure expressing the degree of distrust of an event. Definition: X is a frame of discernment, m: $2^X \longrightarrow [0, 1]$ is given as the basic probability assignment on X. If there are $A \in P(X)$, $B \in A$, then the function Pl: $2^X \longrightarrow [0, 1]$ is defined as

$$\mathrm{Pl}(A) = 1 - \mathrm{Bel}(\overline{A}) = \sum_{B \cap A \neq \Theta} m(B). \tag{13}$$

In the previous equation, $\mathrm{Pl}(A)$ indicates that event A is true uncertainty and $\mathrm{Bel}(\overline{A})$ shows the trust degree of event $\overline{A}$. The degree of mistrust $\mathrm{Pl}(A)$ of A can be determined by equation (13).

The minimum level of trust of evidence theory for event A is $\mathrm{Bel}(A)$, the potential degree of trust in event A is stated as $\mathrm{Pl}(A)$, the support interval of event A is expressed as $[0, \mathrm{Bel}(A)]$, and the likelihood interval of event A is stated as $[0, \mathrm{Pl}(A)]$. When the evidence neither confirms nor denies the occurrence of event A, a trust interval can be used for this uncertain phenomenon, to represent the probability of event A.

4.2. Improvements of D-S Evidence Theory

4.2.1. Evidence-Based Improvements. Modifying the evidence source can reduce the effect of interference factors on the fusion assessment results and improve the evaluation results' accuracy. In this study, the idea of assigning weights is utilized to allocate each evidence's importance to increase the reliability of the evidence on the decision result and weaken the effect of conflicting evidence.

For an uncertain event, there are n pieces of evidence, and the corresponding discernment frame X contains N focal elements with m_i representing the evidence set composed of the basic probability assignment function equivalent to the evidence under each focal element:

$$m_i = \left[m_i(A_1), m_i(A_2), \ldots, m_i(A_n)\right]^T, \quad i = 1, 2, \ldots, n. \tag{14}$$

Equation (14) is utilized to determine the distance between m_i and m_j and d_{ij} represents the distance of m_i and m_j. This distance function with a better reflection in explaining the focal element and the reliability between pieces of evidence can better determine the conflict between pieces of evidence:

$$d_{ij} = d\left(m_i, m_j\right) = \sqrt{\frac{1}{2}\left[\left\|m_i\right\|^2 + \left\|m_j\right\|^2 - 2\left(m_i, m_j\right)\right]}. \tag{15}$$

The similarity function is further derived from equation (15). The similarity between m_i and m_j can be expressed as S_{ij} and the expression of S_{ij} is

$$S_{ij} = 1 - d_{ij}. \tag{16}$$

The smaller the distance between the pieces of evidence, the higher the mutual support. The degree of support for evidence can be stated by the sum of other evidence; then the degree of support for evidence m_i is

$$T(m_i) = \sum_{j=1, j\neq i}^{n} S_{ij}, \quad i = 1, 2, \ldots, n. \tag{17}$$

In this paper, the distance similarity matrix between pieces of evidence is utilized to allocate various weights to each sensor to meet the purpose of modifying the evidence source. To prevent the conservative revised evidence source and losing the advantages of the original evidence, this study adopts retaining the original set of more accurate evidence to guarantee the impact of data fusion. Based on the above ideas and the ratio of the degree of support of the evidence, under retaining a good set of evidence sources, the weight β of the evidence is determined based on the level of support. The specific formula is as follows:

$$\beta(m_i) = \frac{T(m_i)}{\max\left(T(m_i)\right)}. \tag{18}$$

After allocating the weights, the modified basic probability assignment function equivalent to the evidence can be stated as follows:

$$\begin{aligned} m_i'(i) &= \beta(m_i) \cdot m_i, \\ m_i'(\Theta) &= \beta(m_i) \cdot m_i + (1 - \beta(m_i)). \end{aligned} \tag{19}$$

4.2.2. Improvements Based on Fusion Rules. In this paper, using the time series prediction value of the monitoring data of each sensor, the basic probability assignment function value is calculated. After fusing the value of each sensor, the mine gas safety state is judged. The fusion rules of D-S evidence theory are as follows.

According to two independent pieces of evidence M_1 and M_2, the focal elements of the two pieces of evidence are B_i and C_j ($i = 1, 2, 3, \ldots, n$; $j = 1, 2, 3, \ldots, m$), and the basic probability assignment function value after their fusion is $m(A)$:

$$\begin{cases} m(A) = M_1 \oplus M_2 = \dfrac{1}{1 - K} \displaystyle\sum_{B_i \cap C_j = A} m_1(B_i) m_2\left(C_j\right), \\ K(M_1, M_2) = \displaystyle\sum_{B_i \cap C_j = \Theta} m_1(B_i) m_2\left(C_j\right). \end{cases} \tag{20}$$

In the previous equation, $K(M_1, M_2)$ is known as the conflict coefficient representing the degree of conflict between the two pieces of evidence M_1 and M_2. There is no conflict between the two pieces of evidence when the conflict coefficient is 0. However, when it is closer to 1, greater conflict exists between the two pieces of evidence, as a complete conflict.

Many scholars [30–32] believe that the fusion rules of evidence theory are imperfect in the processing of evidence; hence, the reasonable modification of fusion rules can also enhance the accuracy of fusion. After modification of the evidence source, the simple modification of the evidence source data to prevent high conflicts between the pieces of evidence may result in the revised evidence to lose the effective information of the original evidence. The conflict allocation coefficient is introduced based on the fusion rules to enhance the decision stage accuracy.

The conflict allocation coefficient $\omega(A_i)$ can be expressed as

$$\omega(A_i) = \frac{\sum_{i=1}^{n} m_i'\left(A_{ij}\right)}{\sum_{i=1}^{n}\sum_{j=1}^{p} m_i'\left(A_{ij}\right)}. \tag{21}$$

The enhanced formula of D-S evidence theory fusion rule is expressed as

$$m(A) = \sum_{B_i \cap C_j = A} m_1(B_i) m_2(C_j) + K \cdot \omega(A_i). \quad (22)$$

In equation (22), set A denotes the intersection of the focal element B_i and focal element C_j.

4.3. Settings for the Gas Safety Evaluation Model. The gas safety evaluation model includes AWLS-SVM and improved D-S evidence theory, which are explained in detail before. The frame of discernment and the basic probability assignment function are the bases of D-S evidence theory calculation, some settings should be done before to use the evaluation model.

4.3.1. Settings of Discernment Frame. From the perspective of D-S evidence theory, the gas safety state can be considered as a judgmental problem, and the summary of hypothetical results can be explained as a frame of discernment. Based on the coal mine safety regulations [1] and related literature [33, 34], the gas safety state is divided into five states: no danger implies that the working face of the coal mine is in a decent environment; mild danger represents that the working face possesses a certain risk, and this danger value is within the acceptable range, an on-site inspection should be completed; moderate danger implies that the working face is unsafe, the indicated value has exceeded the acceptable range, and an on-site inspection is required as soon as possible; severe danger represents that the working face is very bad, and the staff should be evacuated; and uncertain implies that the evacuated result is vague, and the work should be redone after checking the data source and the evaluation process. Hence, the frame of discernment for the coal mine gas safety evaluation model can be explained as $X = \{X_1$ (no danger), X_2 (mild danger), X_3 (moderate danger), X_4 (severe danger), and Θ (uncertain)$\}$.

4.3.2. Construction of Basic Probability Assignment Function. In this paper, the posterior probability modeling technique is utilized to construct the basic probability assignment function, and the similarity degree is introduced to modify the evidence source. The support degree of each sensor is characterized by the basic probability assignment function to the safety state of mine gas. In this paper, a time series prediction model is made through the AWLS-SVM, and the prediction model is developed with each influence factor as an input to obtain the prediction value of each sensor. The posterior probability modeling technique calculates the basic probability assignment function of each sensor.

Taking a sensor as an example, the basic probability assignment function value obtained by the posterior probability modeling method is y, and the frame of discernment is $X = \{X_1, X_2, X_3, X_4, \Theta\}$. The distance between X and y can be stated as

$$d_i(X_i, y) = |X_i - y|. \quad (23)$$

The correlation coefficient between the evidence and X_i can be stated as

$$c_i = \frac{(1/d_i)}{\sum_{i=1}^{4} (1/d_i)}. \quad (24)$$

Introducing equation (24), the uncertainty $m(\Theta)$ of the corresponding evidence and the basic probability assignment function $m(i)$ can be expressed as

$$m(i) = \frac{c_i}{\sum c_i + E},$$
$$m(\Theta) = \frac{E}{\sum c_i + E}, \quad (25)$$
$$E = \frac{1}{2}|y - x|^2,$$

where y represents the predicted value of the time series prediction model and x shows the expected output value of the prediction model.

5. Case Analysis

5.1. Data Sources. Qing Gang Ping coal mine located in Shaanxi province is taken as the study area, and the data in this paper are obtained from the coal mine monitoring system, which includes the gas concentration at the upper corner (No. A02), the wind speed (No. A09), the gas concentration at the working face 10 meters away (No. A01), the dust (No. A11), the return air tunnel gas concentration (No. A08), and the return air tunnel temperature 15 meters away (No. A07). The original data sampling interval is 1 minute, and the data distribution has obvious jagged characteristics. Hence, this paper uses 5 minutes as the sampling interval to obtain 1500 groups of samples and choose the first 1400 samples for model training and the remaining samples for model testing. The sample set of original monitoring data are shown in Table 1.

5.2. Predicted Results of the Time Series Prediction Model. To predict the monitoring value of each sensor at the next moment, this paper uses the multivariable AWLS-SVM time series prediction model introduced in Section 3. It also uses the target sensor as the output and other sensors as the input for model training. SPSS software was utilized to analyze the Pearson correlation of A02, A01, A09, A11, A07, and A08 monitoring sensors. The analysis results are represented in Table 2.

The interpretation of a correlation coefficient depends on the context and purposes. One of the common criteria used is $|r| > 0.95$, significant degree; $0.8 \le |r| < 0.95$, high degree; $0.5 \le |r| < 0.8$, moderate degree; $0.3 \le |r| < 0.5$, low degree; and $|r| < 0.3$, irrelevant. Thus, the correlation of 0.3 is regarded as the limit in this paper. According to Table 2, the correlation coefficients are all greater than 0.3, and it is

Table 1: The sample set of monitoring data.

Number	A02 (%)	A01 (%)	A09 (m/s)	A11 (mg/m^3)	A07 (°C)	A08 (%)
1	0.224	0.262	1.952	0.02	21.332	0.35
2	0.226	0.26	1.992	0.014	21.3	0.342
3	0.218	0.26	1.97	0.08	21.306	0.342
4	0.218	0.27	1.98	0.082	21.3	0.342
5	0.212	0.276	2.016	0.068	21.304	0.34
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
1497	0.368	0.408	1.926	0.086	22.026	0.502
1498	0.37	0.406	1.916	0.084	22	0.518
1499	0.362	0.4	1.944	0.076	22	0.496
1500	0.352	0.396	1.944	0.074	22	0.482

Table 2: The correlation analysis results of various influencing factors.

	A02	A01	A09	A11	A07	A08
A08	0.572	0.910	0.668	0.324	0.788	1

Table 3: The predicted results of various sensors.

	A02	A01	A09	A11	A07	A08
Predicted results	0.380	0.422	1.912	0.094	22.086	0.504

reasonable for each sensor to be as the input of the target sensor. The prediction results are shown in Table 3.

The prediction results were obtained undertaking the steps described in Section 3.3 using the data sources in Section 5.1 from the coal mine monitoring system, and those values would be the input of the data fusion using the improved D-S theory.

5.3. Experimental Results and Analysis

5.3.1. Contrast Analysis of Conflict Degree. The posterior probability modeling method introduced in Section 4.3 is used in this paper to calculate the basic probability assignment function of each sensor. The BPA of each sensor is shown in Table 4.

According to Table 2, the results of single sensor recognition are A09 $m(X_2) = 0.8079$, A07 $m(X_4) = 0.2399$, A11 $m(X_1) = 0.4939$, A02 $m(X_1) = 0.5551$, A01 $m(X_1) = 0.5664$, and A08 $m(X_1) = 0.5954$. Obviously, there is a great conflict between A09 and A07 and other sensors. Using one sensor evaluation result cannot accurately assess the safety state of coal mine gas. Hence, it is necessary to modify the evidence source before fusion.

This study adopts the improved technique of evidence source introduced in Section 4.2 and redistributes the weights for each sensor based on the BPA in Table 4, and the revised BPA is shown in Table 5.

According to Table 5, A09 is revised from $m(X_2) = 0.8079$ to $m(X_2) = 0.4622$, and A07 is revised from $m(X_4) = 0.2399$ to $m(X_4) = 0.1997$. The conflict is considerably reduced, indicating that the modification method of the evidence source is feasible and retains the excellent evidence of A02. At the same time, Table 5 shows that only using sensors A07 and A09 as evaluation evidence will fail decision-making. Moreover, only using A11, A02, A01, and A08 as evaluation evidence has low recognition accuracy and makes decision reliability low. Therefore, it is not reliable to use only one sensor to assess the safety state of coal mine gas.

Table 4: The basic probability assignment functions.

	A09	A07	A11	A02	A01	A08
X_1	0.0646	0.2057	0.4939	0.5551	0.5664	0.5954
X_2	0.8079	0.2160	0.2358	0.2150	0.2106	0.1979
X_3	0.0557	0.2273	0.1549	0.1333	0.1294	0.1187
X_4	0.0288	0.2399	0.1153	0.0966	0.0934	0.0848
Θ	0.0431	0.1111	0.0000	0.0000	0.0002	0.0032

Table 5: The basic probability assignment function after modifying the evidence source.

	A09	A07	A11	A02	A01	A08
X_1	0.0369	0.1712	0.4914	0.5551	0.5643	0.5801
X_2	0.4622	0.1797	0.2357	0.2150	0.2098	0.1929
X_3	0.0318	0.1892	0.1541	0.1333	0.1298	0.1157
X_4	0.0165	0.1997	0.1148	0.0966	0.0930	0.0826
Θ	0.4526	0.2602	0.0051	0.0000	0.0040	0.0287

5.3.2. Comparative Analysis of Evaluation Results. Based on the comparative analysis of the degree of conflict in Section 5.3, data fusion plays a key role in the decision-making results. For the modifying method of the evidence source, this paper calls the D-S-1 evidence theory. Moreover, the D-S-2 evidence theory is called for the method of modifying the fusion rule. Sensors A09, A07, A11, A02, A01, and A08 are recorded as pieces of evidence e_1, e_2, e_3, e_4, e_5, and e_6. The fusion procedure of multisensors is the fusion process of two sensors in sequence. The comparison outcomes of the multisensor data fusion of the three methods are provided in Figures 2–6.

According to Figure 2, the fusion evidence sources e_1 and e_2 are all highly conflicting pieces of evidence; hence, the decision results of D-S evidence theory and D-S-1 evidence

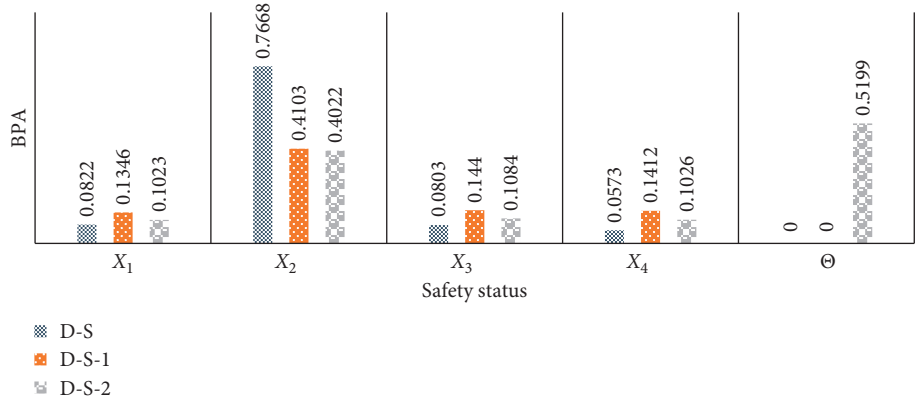


Figure 2: The comparative analysis of e_1e_2 fusion results.

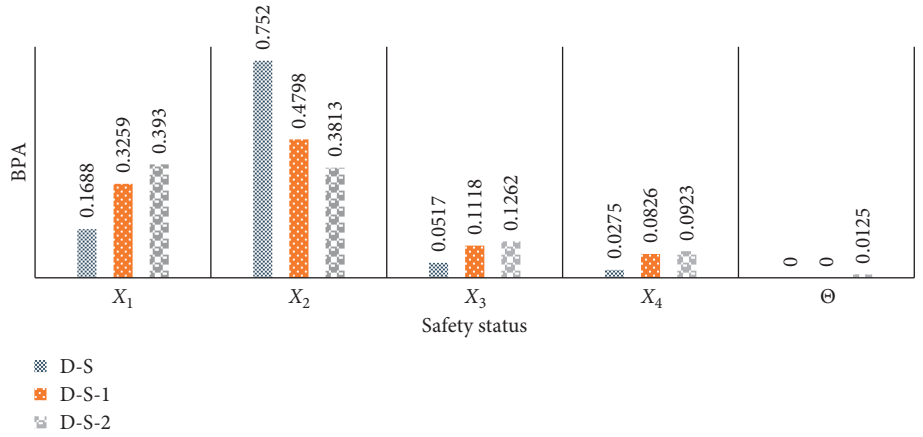


Figure 3: The comparative analysis of $e_1e_2e_3$ fusion results.

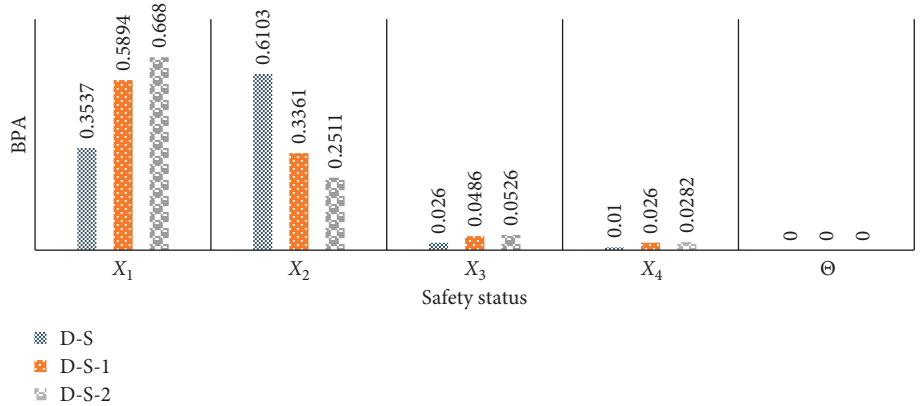


Figure 4: The comparative analysis of $e_1e_2e_3e_4$ fusion results.

theory are invalidated, and the recognition results of D-S-2 evidence theory are uncertain. Followed by introducing the evidence source e_3 in Figure 3, the recognition results of the D-S-1 and D-S evidence theories are wrong, and the D-S-2 evidence theory recognition results are accurate. This proves that the enhanced fusion rule in this paper is effective in retaining the revised evidence source. In Figure 4, according to the fusion results of evidence sources e_1, e_2, e_3, and e_4, D-S evidence theory recognition result is inaccurate and D-S-2 evidence theory recognition results are accurate proving that the modified technique of the evidence source enhanced in this paper is correct, eliminating the interevidence high conflicts. Figures 5 and 6 show that the D-S-2 evidence theory technique for modifying the evidence source and fusion rules in this paper is reasonable. The recognition accuracy of the D-S-2 evidence theory is higher compared to

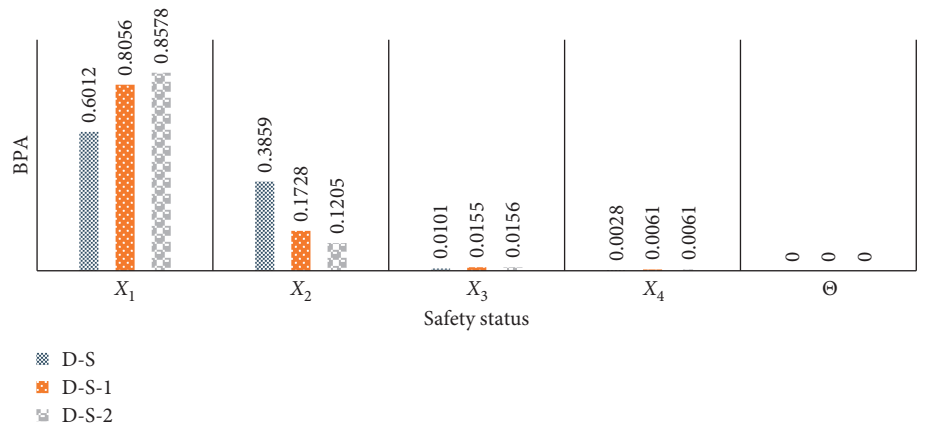


Figure 5: The comparative analysis of $e_1e_2e_3e_4e_5$ fusion results.

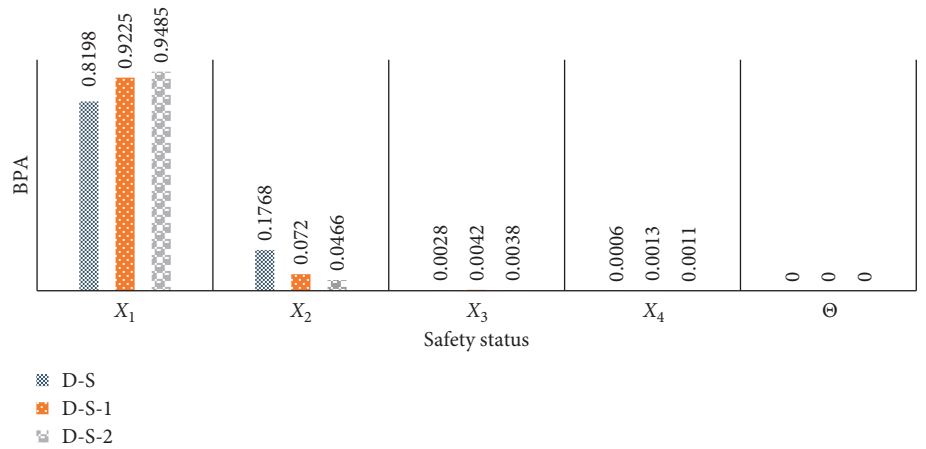


Figure 6: The comparative analysis of $e_1e_2e_3e_4e_5e_6$ fusion results.

the D-S-1 evidence theory and D-S evidence theory. The accuracy of D-S evidence theory based on the improved fusion rules (D-S-2, the model proposed in this paper) is improved by 2.82% (from 0.9225 to 0.9485), respectively, compared to D-S evidence theory based on modified evidence sources (D-S-1) and improved by 15.70% (from 0.8198 to 0.9485) compared to the original D-S evidence theory (D-S).

The accuracy rate of mine gas safety state recognition was enhanced. At the same time, the fusion rule satisfies the exchange law; moreover, it can be concluded that increasing the evidence during the fusion process leads to the higher accuracy of the identification in the decision stage. The problem regarding the difficulty in accurately characterizing the gas safety state in the single sensor is solved. It can be concluded that the multisensor data fusion gas safety state evaluation system suggested in this paper possesses high practical value in field applications with important theoretical significance for overwhelming the occurrence of gas disasters and enhancing the safe and sustainable development of the coal industry.

5.3.3. Model Uncertainty Measure. This paper utilizes Shannon entropy [35] for measuring the uncertainty of the above three D-S evidence theories. Let n signal sources make up the signal $X = \{x_1, x_2, x_3, \ldots, x_n\}$; the probability that each signal source represents the equivalent information for an event is $P = \{p(x_1), p(x_2), p(x_3), \ldots, p(x_n)\}$; then the system structure S of the signal can be stated as

$$S = \begin{pmatrix} X \\ P \end{pmatrix} = \begin{pmatrix} x_1 & x_2 & \ldots & x_n \\ p(x_1) & p(x_2) & \ldots & p(x_n) \end{pmatrix}. \tag{26}$$

Therefore, the Shannon entropy of the signal is given as

$$H(x) = -\sum_{i=1}^{n} p(x_i) \ln p(x_i). \tag{27}$$

The uncertainty of fusion data using D-S evidence theory is

$$\begin{aligned} &-0.8198 * \ln 0.8198 - 0.1768 * \ln 0.1768 - 0.0028 \\ &\quad * \ln 0.0028 - 0.0006 * \ln 0.0006 = 0.4901. \end{aligned} \tag{28}$$

The uncertainty of fusion data using D-S-1 evidence theory is

$$\begin{aligned} &-0.9225 * \ln 0.9225 - 0.0720 * \ln 0.0720 - 0.0042 \\ &\quad * \ln 0.0042 - 0.0013 * \ln 0.0013 = 0.2955. \end{aligned} \tag{29}$$

The uncertainty of fusion data using D-S-2 evidence theory is

$$-0.9485 * \ln 0.9485 - 0.0466 * \ln 0.0466 - 0.0038 * \ln 0.0038 - 0.0011 * \ln 0.0011 = 0.2217. \quad (30)$$

From the above results, it is deduced that the enhanced D-S-2 evidence theory has lower uncertainty compared to the D-S evidence theory and D-S-1 evidence theory and can better assess the safety of coal mine gas.

6. Conclusions

According to the features of coal mine monitoring data, a prediction model is made. By obtaining the predicted values of each sensor, the basic probability assignment function of each sensor is determined to utilize the posterior probability modeling method.

Moreover, a safe assessment model of coal mine gas state is made, and multisensor data fusion is realized. Fusing more sensors, the evaluation results are more accurate. The model in this paper effectively solves the problem of difficulty in accurately characterizing the gas safety state by one sensor.

Furthermore, regarding the problem of evidence fusion failure caused by high-conflict data, this paper represents the similarity for modifying the evidence source of conflict data, which effectively decreases the conflict between the evidence sources. At the same time, to prevent distortion of evidence sources, the conflict assignment coefficients are presented to enhance the fusion rules, and the accuracy of evaluation results is improved. It proves that the enhanced D-S evidence theory can improve accuracy by 15.70% compared to the original D-S evidence theory.

The enhanced method has better generalization ability and higher accuracy for coal mine gas safety evaluation providing a theoretical basis for gas disaster accident prevention. According to the results of coal mine gas safety evaluation, there are some policy implications for coal mine safety in China: although reducing gas in underground coal mine has a positive effect on coal mine safety, it is impossible to completely avoid gas production in the short term. Thus, one available choice is to promote the research of gas monitoring and related safety evaluation technologies and to advocate the use of more efficient and accurate technical means.

Data Availability

The data used to support the findings of this study have been deposited in https://github.com/sun-zhenming/CoalMineMonitoringData.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors gratefully acknowledge the financial support from the National Key Research and Development Program of China (2017YFC0804303) and the Fundamental Research Funds for the Central Universities (no. 2020YQNY10).

References

[1] National Coal Mine Safety Administration, "Coal mine safety regulation," 2020, http://www.chinacoal-safety.gov.cn/xw/zt/lszt/mkaqgc/.

[2] Q. G. Sun, "Current situation of coal mine gas disasters in China and countermeasures," *China Coal*, vol. 40, no. 3, pp. 116–119, 2014.

[3] C. Cai, "Analysis method of gas warning results of coal mine safety monitoring and control system," *Industry and Mine Automation*, vol. 44, pp. 15–18, 2018.

[4] F. Tian, M. Zhang, L. Zhou, H. Zou, A. Wang, and M. Hao, "Qualitative and quantitative differences between common occupational health risk assessment models in typical industries," *Journal of Occupational Health*, vol. 60, no. 5, pp. 337–347, 2018.

[5] P. K. Marhavilas, D. Koulouriotis, and V. Gemeni, "Risk analysis and assessment methodologies in the work sites: on a review, classification and comparative study of the scientific literature of the period 2000–2009," *Journal of Loss Prevention in the Process Industries*, vol. 24, no. 5, pp. 477–523, 2011.

[6] P. K. Marhavilas and D. E. Koulouriotis, "A risk-estimation methodological framework using quantitative assessment techniques and real accidents' data: application in an aluminum extrusion industry," *Journal of Loss Prevention in the Process Industries*, vol. 21, no. 6, pp. 596–603, 2008.

[7] Y. Li and F. W. Guldenmund, "Safety management systems: a broad overview of the literature," *Safety Science*, vol. 103, pp. 94–123, 2018.

[8] A. A. Rassafi, S. S. Ganji, and H. Pourkhani, "Road safety assessment under uncertainty using a multi attribute decision analysis based on Dempster-Shafer theory," *KSCE Journal of Civil Engineering*, vol. 22, no. 8, pp. 3137–3152, 2018.

[9] C.-H. Chen, F.-J. Hwang, and H.-Y. Kung, "Travel time prediction system based on data clustering for waste collection vehicles," *IEICE Transactions on Information and Systems*, vol. E102.D, no. 7, pp. 1374–1383, 2019.

[10] C.-H. Chen, "A cell probe-based method for vehicle speed estimation," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E103.A, no. 1, pp. 265–267, 2020.

[11] C. H. Chen, F. Y. Song, F. J. Hwang, and L. Wu, "A probability density function generator based on neural networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 541, Article ID 123344, 2020.

[12] Y. Chen, C. Gu, C. Shao et al., "An approach using adaptive weighted least squares support vector machines coupled with modified ant lion optimizer for dam deformation prediction," *Mathematical Problems in Engineering*, vol. 2020, 23 pages, Article ID 9434065, 2020.

[13] Q. Zhang, H. G. Li, and H. Li, "An improved least squares SVM with adaptive PSO for the prediction of coal spontaneous combustion," *Mathematical Biosciences and Engineering*, vol. 16, no. 4, pp. 3169–3182, 2019.

[14] S. Zhang, J. Q. Chen, and D. Zhang, "Gob spontaneous combustion prediction based on neural network," *Transducer and Microsystem Technology*, vol. 31, no. 5, pp. 10–12, 2012.

[15] X. Tong, W. Fang, S. Yuan, J. Ma, and Y. Bai, "Application of Bayesian approach to the assessment of mine gas explosion," *Journal of Loss Prevention in the Process Industries*, vol. 54, pp. 238–245, 2018.

[16] L. M. Pejic, J. G. Torrent, E. Querol, and K. Lebecki, "A new simple methodology for evaluation of explosion risk in underground coal mines," *Journal of Loss Prevention in the Process Industries*, vol. 26, pp. 1524–1529, 2013.

[17] M. Li, H. Wang, D. Wang, Z. Shao, and S. He, "Risk assessment of gas explosion in coal mines based on fuzzy AHP and Bayesian network," *Process Safety and Environmental Protection*, vol. 135, pp. 207–218, 2020.

[18] Q. Zhang, "Risk assessment of gas explosion disaster based on random forest model," *IOP Conference Series: Earth and Environmental Science*, vol. 446, Article ID 22081, 2020.

[19] X. D. Sun, L. X. Zhang, and H. L. Qi, "Coal mine safety production risk assessment model based on fuzzy analytic hierarchy process," *Industrial Safety and Environmental Protection*, vol. 40, no. 1, pp. 65–68, 2014.

[20] Z. Dai, J. H. Liu, and W. Wang, "Collection and fusion of gas density data and its security warning for coal mine," *Computer Engineering and Design*, vol. 37, pp. 783–787, 2016.

[21] H.-g. Peng, J.-q. Wang, and P.-f. Cheng, "A linguistic intuitionistic multi-criteria decision-making method based on the Frank Heronian mean operator and its application in evaluating coal mine safety," *International Journal of Machine Learning and Cybernetics*, vol. 9, no. 6, pp. 1053–1068, 2018.

[22] Q. Wang, H. Wang, and Z. Qi, "An application of nonlinear fuzzy analytic hierarchy process in safety evaluation of coal mine," *Safety Science*, vol. 86, pp. 78–87, 2016.

[23] R. J. He and Y. H. Du, "Application of improved particle-swarm-optimization neural network in coalmine safety evaluation," *IOP Conference Series: Materials Science and Engineering*, vol. 423, Article ID 12007, 2018.

[24] X. Li, N. W. Li, and Z. Yang, "Coal mine safety evaluation model based on quantum genetic algorithm," *Computer Systems and Applications*, vol. 21, no. 7, pp. 101–105, 2012.

[25] J. A. Zhang, W. J. Li, and Y. L. Guan, "Application of improved FNN in coal mine safety production warning system," *Coal Engineering*, vol. 8, pp. 168–171, 2013.

[26] D. Wang, L. Liu, and X. M. Zhang, "The improvement and application of the grey correlation degree method in the evaluation of coal mine intrinsic safety," *Journal of Safety Science and Technology*, vol. 9, no. 1, pp. 151–154, 2013.

[27] D. W. Li, S. Li, and M. J. You, "Research on mine safety situation prediction model: the case of gas risk," in *Proceedings of the 2019 11th International Conference on Wireless Communications and Signal Processing (WCSP)*, pp. 1–6, Xi'an, China, October 2019.

[28] J. A. K. Suykens, J. De Brabanter, L. Lukas, and J. Vandewalle, "Weighted least squares support vector machines: robustness and sparse approximation," *Neurocomputing*, vol. 48, no. 1–4, pp. 85–105, 2002.

[29] D. J. Cummins and C. W. Andrews, "Iteratively reweighted partial least squares: a performance analysis by Monte Carlo simulation," *Journal of Chemometrics*, vol. 9, no. 6, pp. 489–507, 1995.

[30] L. Zadeh, "A simple view of the Dempster–Shafer theory of evidence and its implication for the rule of combination," *AI Magazine*, vol. 7, no. 2, pp. 85–90, 1986.

[31] J. An, M. Hu, L. Fu, and J. Zhan, "A novel fuzzy approach for combining uncertain conflict evidences in the Dempster-Shafer theory," *IEEE Access*, vol. 7, pp. 7481–7501, 2019.

[32] J. Schubert, "Conflict management in Dempster-Shafer theory using the degree of falsity," *International Journal of Approximate Reasoning*, vol. 52, no. 3, pp. 449–460, 2011.

[33] H. T. Liu and L. L. Li, "Comprehensive evaluation analysis of mine gas safety based on integrated method," in *Proceedings of the 2009 International Conference on Computational Intelligence and Software Engineering*, pp. 1–4, Wuhan, China, December 2009.

[34] Q. J. Qi, X. L. Zhao, and B. C. Song, "Pre-evaluation method of coal mine safety based on continental distance model with varying weight," *Procedia Earth and Planetary Science*, vol. 1, no. 1, pp. 180–185, 2009.

[35] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 4, pp. 623–656, 1948.

Hindawi
Discrete Dynamics in Nature and Society
Volume 2020, Article ID 1843697, 19 pages
https://doi.org/10.1155/2020/1843697

Research Article

An Integrated DEMATEL-ANP Approach for Mobile Banking Adoptions in China Market

Xiaoning Zhu[1,2] **Rui Yan**[3] **and Tian xing Xu**[1]

[1]*Donlinks School of Economics and Management, University of Science and Technology Beijing, Beijing 100083, China*
[2]*Beijing Low-carbon Operations Strategy Research Center, University of Science & Technology, Beijing 100083, China*
[3]*School of Management and Economics, Beijing Institute of Technology, Beijing Information Science and Technology University, Beijing 100083, China*

Correspondence should be addressed to Rui Yan; yanrui@ustb.edu.cn

Received 15 June 2020; Revised 10 August 2020; Accepted 7 September 2020; Published 10 October 2020

Academic Editor: Fuqiang Gu

With the rapid development of the mobile device and Internet usage, the mobile banking has seen remarkable growth across the world, especially in the past few years, as one of the emerging financial innovations. This paper attempts to explore the mobile banking services offered by selected commercial Chinese banks. The main data of this study were obtained through the comprehensive evaluation of mobile banking services. The DEMATEL-ANP method was used to evaluate the main factors influencing the adoption of mobile banking. The results show that there is still big space for the improvement of the mobile banking services for the selected Chinese banks. The management significance and suggestions for these banks are also discussed.

1. Introduction

Nowadays, with the development of wireless technology, mobile banking (MB) has become one of the most popular ways of dealing with financial service [1]. According to Laukkanen and Cruz [2], MB can make a client independently access banking services without temporal and spatial restrictions. Thanks to this promotion in MB, telecommunication technology and wireless breakthrough in the mobile have made significant progress in this system. Another major factor behind MB development is identified as the technology to provide people with high quality services over more innovative and cost-effective channels to meet client needs [3].

China is one of the fastest mobile and telecommunication technology growing countries in Asian; and there are three mobile service providers working in the China market. By the end of 2016, more than 900 million clients used MB in the Chinese banking market [4]. China's banks are mainly divided into three major policy banks, five state-owned banks, and other commercial banks. Luarn and Lin [5] investigated whether it is a state-owned bank or a commercial bank in China which is progressively introducing MB systems to reduce costs and enhance productivity. Meanwhile, both traditional banks' branches and customers benefit from MB. The banks can save operating costs, while for customers, the convenient operating rules and more flexible use of time can save their effort and time through using MB [6]. In addition, the use of MB also provides impetus to economy for the whole country development. For example, the convenience and security of transaction contributes to the improvement of the whole economy [7]. Consequently, MB is an innovative application to handle bank transactions by which customers interact with a bank via a mobile device. In the same vein, MB is an indispensable part in personal financial business and commercial arena.

However, it must be noted that consumer adoption and usage of new technology is significantly influenced and driven by the speed and ease of connecting with service providers [8]. Most customers have less interest in this service of MB and the adoption rates also do not reach the desired level in developing countries [9]. For instance, Alalwan et al. adopted nearly 4000 clients' information of

mobile services around the world, but only 19% of the customers are using MB services. Because MB is still in its infancy stage of development, leaving plenty of room to improve, to increase the adoption rate, banks need to understand customers' needs and find the factors influencing their thoughts to adopt MB services. Meanwhile, Friedman et al. [10] conducted MB research and revealed that the factors of perceived trust or credibility of users, in relation to web systems, play an important role to use MB, in online shopping, and to communicate sensitive personal information. Because of the unexpected disclosure of information, customers refuse to provide sensitive personal information to banks. The lack of trust makes customers more worried that their information will be revealed to the third party without their consent. The prior research has tried to explain the adoption factors, such as technology advances, better understanding of technology, and more tech-savvy consumers [11]. In the meantime, many researchers have given greater emphasis to perceived risk as a precondition of behavioral intention [12, 13]; the extent research is mostly conducted in accordance with eight aspects, such as performance, financial, time, psychological, social, privacy, physical, and overall risk [14], but most articles have inadequately considered the factors; the comprehensiveness and depth of these researches are not enough at the same time. In this paper, we consider nine factors to study the degree of adoption in MB, such as, design, usefulness, popularity, content, security, common services, convenience services, finance, and livelihood. In addition, this related information can help MB service providers to design corresponding systems that are more responsive to customers.

In short, this paper focused on identifying the main factors influencing the adoption of MB. The remainder of this paper is organized as follows. Section 2 introduces the related literature about MB and factors influencing adoptions. Following is a brief introduction about the ANP-DEMATEL (analytic network process-decision-making trial and evaluation laboratory) method used in this research. Section 4 describes an empirical analysis of evaluating the MB in China. The basic data derives from a comprehensive of MB evaluation. In the end, the purpose of this article is to examine and empirically identify the factors that affect the user's adoption and intention of MB in the Chinese banking context.

2. Literature Review

As one of the most convenient and serviceable technologies, MB has the potential to offer both banking and transaction services to clients in the development of market [15]. Due to cost-effectiveness and greater reachability, banks prefer the MB channel and encourage customers to adopt m-banking services. In the context of MB, consumers should have accession to the required resources. Most consumers visit a bank branch because they do not have access to MB or they are not aware of it [16]. Hence, several efforts were made to explore factors affecting m-banking adoption and usage behavior in the previous literature [17–19]. Changchit et al. [19] collected 309 subjects and used multiple regression analysis to analyze data in this study. Various studies related to MB adoption have been conducted. For instance, a plethora of prior studies, such as Luo et al. [20] and Lee et al. [21] as well as Laukkanen and Cruz [2] have investigated the key factor of risk perception in the original adoption stage of wireless Internet platform, and the results can be used to better understand and improve the acceptance of this specific case for mobile communication. Moreover, numerous studies indicate that intention, performance expectancy, ease of use, compatibility, and self-efficacy can also be used in the research of MB [22–24]. In the extended domain of MB, previous researches have identified the attitude of clients. For instance, Sahoo and Pillai [23] find that attitudes towards banking mediate and MB service scape may affect the customer participation. In addition, Kuisma et al. [24] as well as Bhatiasevi [25] compared the customer-perceived value in the Internet and mobile bill paying service. In the meantime, satisfaction of the users and experience level have been studied. For example, Lin et al. [26] conceptualized their model and found that there are some similarities and differences between high and low experience groups in terms of the MB quality.

Both internal and external perceived behavioral control factors were first differentiated by Ajzen and Madden [27]. It was then increasingly adopted by analyzing the factors that influence adoption of MB in the world. Various studies have been carried on the control factors of behavior. For instance, Alalwan et al. [9], using the Unified Theory of Acceptance and Use of Technology methods, found that the performance of the behavioral intention and motive of hedonic price are greatly affected by the value and risk perception. Based on survey data and descriptive data analysis, Alalwan et al. [28] examined the factors of bank customer intention and behavior to influence the adoption of MB.

Usefulness is one of the main factors that affect the use of MB [29]. Both Alalwan et al. [9] and Hanafizadeh et al. [15] investigated the importance of usefulness on adoption behavior which can best explain the key factors to the Jordanian customers' intention to use MB. In order to better understand the adoption behavior, Alalwan et al. [9] applied the above research to propose the conceptual model and relevant factors, and the major barriers are unable to capitalize upon the usefulness [30, 31].

Demographic factors, namely, gender and experience, are dominant dimension of the traditional service in MB, such as a study by Riquelme and Rios [30] which examined the gender in the adoption of MB effect and the influence of MB adoption in Singapore. Alalwan et al. [9] also investigated the intention of MB adoption in Jordan. Main statistics (mean and standard deviation) with the intention of adoption show that the fluctuation comes from the difference between the customer demographics and intention of MB adoption.

Several researchers have investigated the MB adoption behavior using both qualitative and quantitative methods [1, 17, 18, 31, 32]. The technology acceptance model (TAM) [16], innovation diffusion theory (IDT) [33], unified theory of acceptance and use of technology (UTAUT I and II) [28],

and theory of planned behavior (TPB) [34] were the models used in previous studies to explore m-banking adoption behavior. Literature suggests that perceived ease of use, perceived usefulness, trust, convenience, risk, perceived behavioral control, compatibility, and facilitating conditions are the major factors affecting MB adoption intention [17].

Trust is considered related to, but different from, other predictors. This refers to the extent to which a person believes in MB. Lin [33] and Zhou [35] as well as Kim et al. [36] verify user's initial reliability in MB. They found that trust has a stronger influence on the adoption behavior compared to other factors. In addition, Luarn and Lin [5] proposed the extension work of MB based on the trust construct and two resource-based constructs, and on this basis, Zhou et al. [37] suggest that structural assurance and information quality are key factors affecting initial confidence. The summary of relevant influencing factors is shown in Table 1.

The overall evaluation of MB both in theory and in practice has proven to be very important and complex. However, the research on MB is not enough and there have been limited researches in the current literature. This paper extends the study of these factors discussed above and strives to contribute to subsequent research.

3. Research Methodology

The research on causal analysis has become more and more popular in recent years. AHP method, BP neural network, fuzzy evaluation theory of application, and the establishment of mathematical model can make causal analysis results more objective and fair. Along with fuzzy AHP and Internet banking, "customer service" is an important factor affecting MB efficiency[26]. As with other types of technology acceptance models (TAM), Luarn and Lin [5] added the trust construct (perceived credibility) and two resource-based constructs (perceived self-efficacy and perceived financial cost) of determining factors of the MB user authorization model, and on this basis, Kim et al. [36] integrated TAMs and initial trust to research the adopted information. With the method of unified theory of acceptance and use of technology (UTAUT), Martins et al. [43] described the risk of internet banking based on perception and action. However, Zhou et al. [42] focused on explaining user adoption of technology perceptions. In the meantime, Gu et al. [38] used structural equation modeling (SEM) to test the causal relationship and verify the behavioral intention of MB. However, these methods ignore an influence or feedback relationship between evaluation indicators.

Bayazit and Karpak [44] as well as Karpak and Topcu [45] proposed the selection of evaluation indexes and evaluation methods of appropriateness, and the choice of the characteristics of the evaluation index and evaluation method has a certain relevance. The DEMATEL method, also known as the decision-making trial and evaluation laboratory, is the logical relationship between elements by direct effect on the matrix structure, suitable for the study of complex systems of the interaction relationship between factors [46, 47]. ANP is an extension of AHP (analytic hierarchy process), which allows for quantifiable or unquantifiable coexistence of multiple indicators, which is suitable for a system with internal dependency and feedback relationship [48]. Using the DEMATEL method to evaluate the influence factors of MB, although it can handle the causal relationship between the indexes, the calculation of the index weight is the same and does not take into account the weight of high and low points [49]. As a result, the DEMATEL method is used to obtain impact factors evaluation matrix, and the overall impact of the ANP method calculated the index weight; it is concluded that the mixed weight of indicators, which reflects the relationship between the factors of size and considering the factor weight, make up for the deficiency of the single-use DEMATEL method. An overview of the integrated DEMATEL and ANP approach is given in Figure 1.

Briefly, the values are gathered with questionnaires. In cases of inner dependency, DEMATEL is employed. This method can effectively identify the causal relationship and the degree of interaction between the criteria or elements. Further calculations are carried out with ANP. ANP is to demonstrate the complexity of the system network structure; network of each element may influence each other constraints; the characteristics of the network structure happen to be a reasonable description of the interaction relationship between real things. The integrated DEMATEL and ANP methodology can be summarized in the following steps [50–53]:

Step 1: define the element and judge the relationship: analyze and define the elements in the system and judge the relationship between the two elements according to the expert discussion and questionnaire.

Step 2: construct direct relation matrix: first, the relationship between elements is expressed in the following three influences [50]: no influence (0), low influence (1), medium influence (2), high influence (3), and very high influence (4). DMs require experts to identify the degree of comparison. If there are n elements, then we can get an $n \times n$ matrix based on comparison results, which is called the direct relationship between the matrix.. Here, a_{ij} stands for the degree to which the criteria i affects the criteria j.

Step 3: normalize the direct relation matrix: by multiplying the elements of the direct relationship matrix by λ, the normalized direct relation matrix can be obtained, using the following formulas:

$$M = \lambda * A, \tag{1}$$

$$\lambda = \min\left(\frac{1}{\max\sum_{1\le i\le n}^{n} j = 1\left|\alpha_{ij}\right|}, \frac{1}{\max\sum_{1\le i\le n}^{n} i = 1\left|\alpha_{ij}\right|}\right), \quad i, je\{1, 2, 3, \ldots, n\}. \tag{2}$$

Table 1: Influencing factors of customer satisfaction.

	Influencing factors of customer satisfaction									
Factor Authors	Trust	Cost-effective	Self-efficacy	Risk perception	Behavioral intention	Ease of use	Performance expectation	Compatibility	Value	Usefulness
[3]		√								
[5]		√					√			
[5]	√	√	√						√	
[36]	√	√		√					√	
[38]	√				√	√				√
[26]	√				√	√		√		
[2]				√					√	√
[39]					√		√			
[2]			√						√	
[25]	√				√		√			
[21]				√	√				√	
[40]	√			√	√		√			
[41]			√	√		√				√
[20]	√			√						
[15]	√	√		√		√		√		√
[42]	√									
[36]	√						√			

Step 4: calculate the total-relation matrix: after obtaining the normalized direct relation matrix X, because of $\lim_{k\longrightarrow\infty} X^k = 0$, the comprehensive relational matrix T can be obtained from the following formula, where 0 is the zero matrix and I is the identity matrix. Calculation formula (3) of the integrated influence matrix is as follows:

$$T = N + N^2 + N^3 + \cdots = \sum_{I=1}^{\infty\leftarrow} N^i = N(I-N)^{-1}. \tag{3}$$

Step 5: calculate the influence degree and the influence degree of the calculation elements, such ask_{ij} $(i, j = 1, 2, \ldots, n)$; j is affected by i, as indicated in formulas (4)–(6). The sum of the rows (D_i) is denoted as the comprehensive influence value of the corresponding elements which is called the influence degree, and the corresponding elements of each column (R_j) are affected by the comprehensive influence of other elements, which is called the influenced degree.

$$K = \left[K_{ij}\right]_{n*n}, \quad i, j \in \{1, 2, 3, \ldots, n\}, \tag{4}$$

$$D_i = \sum_{i=1}^{n} K_{ij}, \quad (i = 1, 2, \ldots, n), \tag{5}$$

$$R_j = \sum_{j=1}^{n} K_{ij}, \quad (j = 1, 2, \ldots, n). \tag{6}$$

Step 6: calculate a threshold value $(D+R, D-R)$. $D+R$ are the elements in the evaluation index system of the size, and the higher the degree, the more the important it is. $D-R$ is the reason factor, and the element is greater than 0, which indicates that this factor has influence on other indexes. If less than 0, this factor means a lot of room to improve.

Step 7: draw a causal diagram. By mapping the dataset of the $(D+R, D-R)$ and using the horizontal axis indicating $D+R$ and the vertical axis indicating $D-R$, the impact-diagraph map is obtained. Results analysis can clearly identify the relationship between indicators and its importance. In the meantime, the result can find out the main factors that influence the adoption of MB.

Step 8: construct the network of the considered problem and evaluate the remaining nodes and alternatives using the ANP [54].

Step 9: construct the network structure and two comparison judgment matrices. As the correlation index of the degree is smaller, when calculating the relative weight, the impact is not large; so, by setting the threshold value (threshold) to remove the smaller correlation connections, using the DEMATEL questionnaire safety evaluation index of comprehensive influence relation matrix T, we can remove less than the threshold value of the element. Then, we can draw the system network architecture according to the remaining relation. According to the expert discussion results and ANP evaluation scale table (as shown in Table 2), the two comparison judgment matrices between each element can be formed. Therefore, the transpose judgment consist of the reciprocal of these values.

Step 10: calculate the eigenvalues and eigenvectors of the comparison matrix. Supposing that the general goal of the control layer is A, its control criteria are $(C_1, \ldots, C_i, \ldots, C_n)$and the pairwise comparison matrix $A = a_{ij}$. The column vectors in this matrix represent the influence of all the elements inC_i on each element in C_j. Saty introduced the row vector average method and normalized the results with formula (7) as follows:

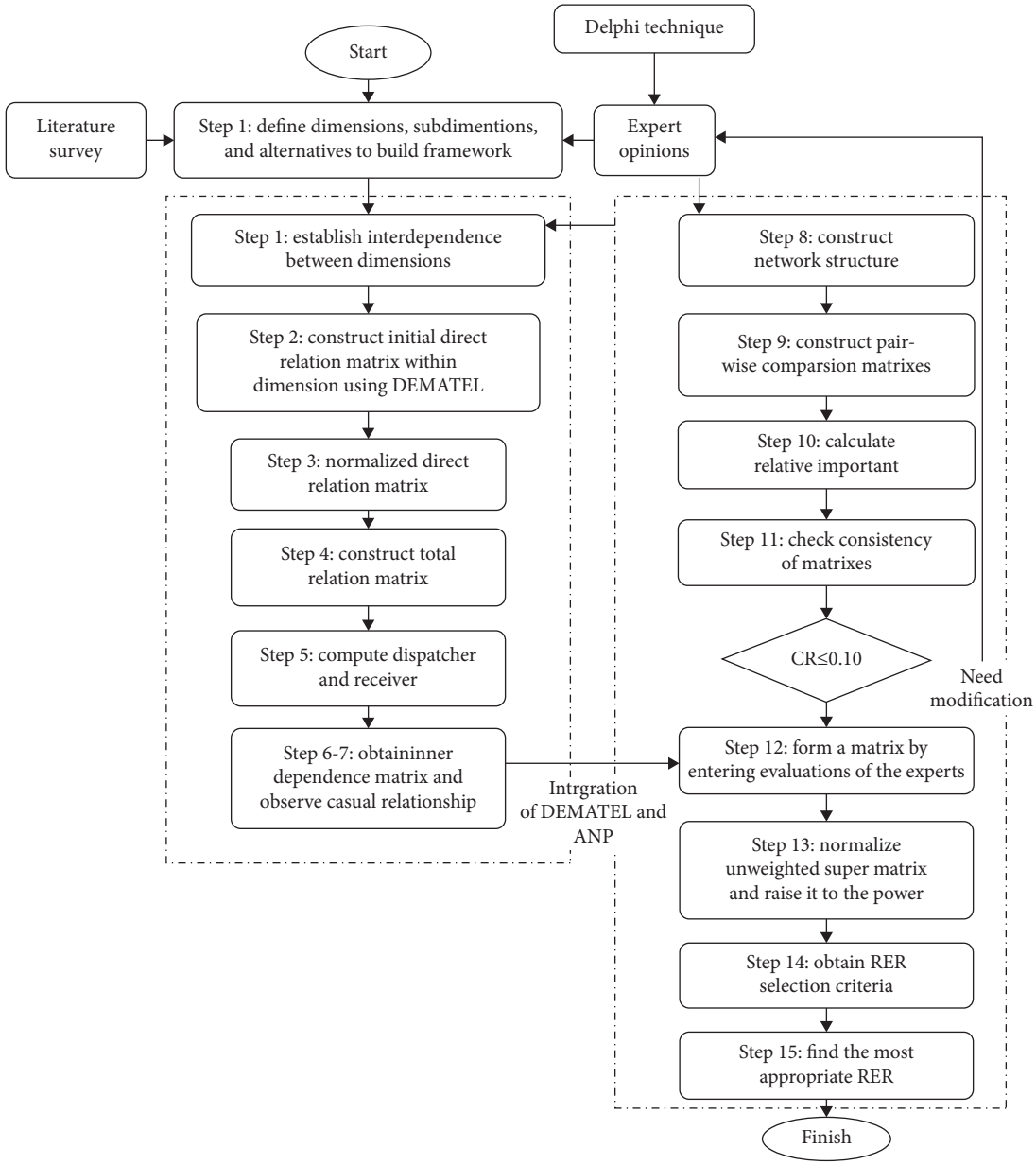


Figure 1: Overview of the integrated DEMATEL and ANP approach.

$$W_i = \frac{\sum_{j=1}^{n}\left(a_{ij}/\sum_{i=1}^{n} a_{ij}\right)}{n}, \quad \forall i, j = 1, 2, \ldots, n. \tag{7}$$

Here, the comparison matrix A completely responds to $a_{ik} = a_{ij}.a_{jk}$, $\forall i, j, k$. The largest eigenvalue $\lambda_{\max}$ can be concluded in the following formula:

$$AW = \lambda W \lambda_{\max} = \frac{1}{n}\sum_{i=1}^{n}\frac{(AW)_i}{W_i}. \tag{8}$$

Step 11: check consistency test: based on formulas (9) and (10), the consistency test was performed. CR represents the consistency ratio; CI represents the consistency index, and RI is a random indicator.

$$\text{C.I} = \frac{\lambda_{\max} - n}{n-1}, \tag{9}$$

$$\text{C.R} = \frac{\text{C.I}}{\text{R.I}}. \tag{10}$$

If the CR is less than 0.1, the pairwise comparisons are acceptable; otherwise, they are not acceptable. RI's control index is shown in Table 3.

Step 12: after complete consistency check, we can get the unweightd supermatrix by integrating the index of characteristic vector into the supermatrix, and unweighted supermatrix is multibplied by $\lambda_{\max}$ to determin the weighed supermatrix. Then, calculate the "limit supermatrix" of the matrix, the matrix of each

Table 2: ANP evaluation scale table.

Important degree	Definition	Explanation
1	Equal importance	The contribution of the two indicators is of equal importance
3	Moderate importance	Experience judgment shows a slight approval of an indicator
5	Strong importance	Empirical judgment strongly endorses an indicator
7	Very strong importance	The facts show that they strongly agree with an indicator
9	Extreme importance	There is enough evidence to support a certain indicator
2, 4, 6, 8	Intermediate values	Between the above evaluation criteria

Table 3: RI's control index.

Matrix order	1	2	3	4	5	6	7	8	9	10
RI	0	0	0.52	0.89	1.11	1.25	1.35	1.4	1.45	1.49

element is the various elements of the weights in the network, so as to get the relative order of all elements in the system [55].

Step 13: determine the most suitable alternative: the weights are calculated for ranking the alternatives. Using the supermatrix, the alternative that has the highest overall priority value is selected.

4. Data Collection and Results Analysis

Due to the rapid growth of China's macro economy and the stable monetary policy, the financial deleveraging has achieved remarkable results. China's banking sector has picked up and the nonperforming loan ratio has fallen (2017 China banking development trend report). According to the latest CBRC official report, the number of commercial banks has reached thirty thousand, and the number is gradually expanding. The details are as follows: banks in China, including three policy banks, five large commercial banks, one postal savings banks, twelve share-holding commercial banks, one hundred and thirty three urban commercial banks, five private banks, eight hundred and fifty nine rural commercial banks, seventy one rural cooperative banks, one thousand three hundred and seventy three rural credit cooperatives, one home DE housing savings banks, one thousand three hundred and eleven rural banks, and forty eight rural capital supports, a total of three thousand eight hundred and twenty two, as shown in Table 4. Considering the differences between large commercial banks and small rural credit cooperatives in terms of business resources and building MB capabilities, ten commercial banks (five state-owned commercial banks and five national joint-stock commercial banks) were selected for this study.

Four raters are participated to make an appropriate rating and each of them needs to evaluate all selected ten banks in China. Based on a pretest with selected criteria for consistency consideration, the selected items from all types of mobile banking are rated with the commonly used Little Scale, i.e., from a scale of 1 (being the worst possible scenario) to 5 (meaning excellent) accordingly. The average value is taken as the final evaluation result.

Most of the banks in China can achieve customers' satisfaction in terms of visual effect. For example, considering the visual effect, only 10% of observed banks scored less than three points; 20% of observed banks need to improve their structure classification; 20% of the Chinese banks are lacking visual appeal, as shown in Table 5. Regarding to the visual effect of the MB, we evaluate the color assortment, site aesthetic, and visual attraction. All of the Chinese banks that were observed used texts and pictures to introduce their product information, but none of the banks had video clips, but the video was more convincing to customers. In short, more than 80% of the banks can meet the needs of customers in the MB design, but there is still a lot of room for improvement.

The size of the MB, number of downloads, convenience of the research, version, and replacement of the page are all important factors that affect the users' adoption of MB. The convenience of use is a very important thing for customers, if the steps are too complex, it will be possible to forgo the use directly. Since the user's mobile phone memory is limited, the size of the software becomes the primary factor affecting the user's installation of MB. According to the results of the investigation, the memory of the 60% of observed banks is concentrated in 30 M– 60 M, and the other 40% banks have two less than 30 M and two more than 60 M, with the largest memory footprint of 87.9 M, as shown in Table 6. Similarly, the number of downloads that MB needs is also a key factor affecting the user's use. Research shows that 80% of selected banks only need to download one software to make it all work. All banks can meet the needs of customers in their search for convenience. At the same time, MB also includes both IOS and Android versions. In addition, only one bank needs to relog on the page when changing to another interface, and the other nine selected banks can switch pages more easily.

The collected data in terms of information content and the security of MB are summarized in Table 3. According to the observation, all the Chinese banks under investigation are required to verify their online payments, for example, the five methods of verification and two approaches of testing each account for 10%; 20% of the observed banks have four ways and a large portion (60%) of Chinese banks have three methods to verify it, as shown in Table 7. However, most observed banks are unsatisfactory in ensuring safe innovation, such as facial and fingerprint unlocking, with only 30% of banks doing so. It can be assumed that protecting customer's information security is the premise of gaining users' trust, and 90% of the banks can do better to protect customer information. For example, in the page operation of some MBs, it is necessary to have corresponding permission to carry out; on the basis of no security measures, private client information is not allowed on the Internet and so on.

Table 4: The list of the selected commercial banks in China.

Country	No.	Commercial bank	Type of the bank
China	B1	Industrial and Commercial Bank of China	State-owned commercial bank
	B2	Construction Bank of China	State-owned commercial bank
	B3	Bank of China	State-owned commercial bank
	B4	Agricultural Bank of China	State-owned commercial bank
	B5	Bank of Communications	State-owned commercial bank
	B6	China Postal Savings Bank	National joint-stock commercial bank
	B7	China Everbright Bank	National joint-stock commercial bank
	B8	China Merchants Bank	National joint-stock commercial bank
	B9	China Minsheng Bank	National joint-stock commercial bank
	B10	Bank of Beijing	National joint-stock commercial bank

Table 5: Summary of the visual effect of the MB.

Visual effect of the MB			
Category	Ranking	Number	%
Color assortment	Need improvement (≤3)	1	10
	Satisfied (>3)	9	90
Total		10	100
Site aesthetic	Need improvement (≤3)	2	20
	Satisfied (>3)	8	80
Total		10	100
Visual attraction	Need improvement (≤3)	2	20
	Satisfied (>3)	8	80
Total		10	100

Table 6: Summary of the ease of use.

Ease of use			
Category	Ranking	Number	%
File size	(0, 30 M]	2	20
	(30 M, 60 M]	6	60
	(60 M, 100 M]	2	20
Total		10	100
Number of APPs to download	Number (=1)	8	80
	Number (=2)	2	20
Total		10	100
Finding convenience	Need improvement (≤3)	0	0
	Satisfied (>3)	10	100
Total		10	100
Version	IOS	10	100
	Android	10	100
Problems with replacing the page	Replace the page (1)	1	10
	Do not replace the page (0)	9	90
Total		10	100

The information of common services is very important for banks, especially for the MB. In this research, all the banks under investigation can meet customers' requirement about account management. Over half (8 out of 10) of the observed banks can apply for credit card through MB, which implies that this method eliminates the need to handle long waits at the counter and reduce banking business. Surprisingly, 60% of the Chinese banks need to improve the management of deposit and loan. However, this server can not only make customers more convenient to handle deposit business but also can increase the bank's business performance. With the improvement of the living standard in our country, the annual number of tourists traveling abroad increases year by year, and the demand for foreign currency is gradually increasing. Nevertheless, half of the MB do not have this business, and only 20% of observed banks manage well in foreign currency management, as shown in Table 8. Since most banks need to apply for foreign currency in advance, clients have to wait a certain amount of time to get the corresponding foreign currency, especially the small currency. Therefore, it is necessary for banks to increase this business, not only to improve the work efficiency of the outlets but also to some extent to attract potential users.

Most of the convenience services, such as mobile payment, no card withdrawals, foreign currency cash reservation, cross-border payments, and network reservation, can

Table 7: Summary of security.

Security			
Category	Ranking	Number	%
Online payment or remittance security method	Method (=5)	1	10
	Method (=4)	2	20
	Method (=3)	6	60
	Method (=2)	1	10
	Method (=1)	0	0
Total		10	100
Client confidentiality	Need improvement (≤3)	1	10
	Satisfied (>3)	9	90
Total		10	100
Ensure safe innovation (fingerprint, face)	Yes (1)	3	30
	No (0)	7	70
Total		10	100

Table 8: Summary of the available common services.

Ease of available common services			
Category	Ranking	Number	%
Account management (balances, details)	Need improvement (≤3)	0	0
	Satisfied (>3)	10	100
Total		10	100
Transfer money	Need improvement (≤3)	0	0
	Satisfied (>3)	10	100
Total		10	100
Apply for a credit card	Need improvement (≤3)	2	20
	Satisfied (>3)	8	80
Total		10	100
Deposit and loan management	Need improvement (≤3)	6	60
	Satisfied (>3)	4	40
Total		10	100
Foreign currency management	Need improvement (≤3)	8	80
	Satisfied (>3)	2	20
Total		10	100

bring convenience to customers. Mobile payment is widely used in both large and corner stores in China. According to our observation, 30% of the observed banks need to improve their mobile payment method. Similarly, a large portion of MB needs to improve the no card withdrawals as well as the foreign currency cash reservation services, and even more surprisingly, only 10% of the selected banks provide the cross-border payments services on their MB, as shown in Table 9. In addition, due to the limited bank outlets, queuing users will take a long time to do their business, this is not only a waste of time for the user but also reflects that the bank efficiency is low from the side, and cannot satisfy the needs of customers.

Except the convenience services, more and more customers start concentrating on the investment/finance management, so we collect the data in this research which are summarized in Table 10. According to our survey, all the observed banks offer the investment/finance management services. However, 10% of the selected banks need to improve the extensive products and 20% of the banks need to improve the detailed introduction to customers. Otherwise, these imperfect services may cost a long time for clients to find useful information.

The livelihood payment is the addition items to help people settle the cost of living, so in large majority literatures the livelihood payment is identified as the significant index for evaluating the payment method. In this research, among the 10 banks surveyed, there are options for people's livelihood, but the details of the content and information are different. For example, as shown in Table 11, 90% of observed banks have the functions of charging, Internet, and electricity; 60% of the banks can offer food coupons and movie tickets and 70% of the banks have the option of paying a fine or purchase a ticket on MB. In addition, 100% of mobile phone banks can be registered by the corresponding option, but 70% of the banks have fewer than 10 hospitals to choose from, and the largest number of hospitals have 30. However, most banks do not provide the corresponding service of health management, only 10% of the bank provide this service, but the content is not so detailed.

The number of downloads and user comments are key factors that affect the popularity of MB. As can be seen from Table 8, half of the MB have been downloaded from 0–50 million times, and 30% of the banks are concentrated in 50–100 million times, with the maximum number of

Table 9: Summary of the available convenience services.

Ease of available common services			
Category	Ranking	Number	%
Mobile payment (collection, payment)	Need improvement (≤3)	3	30
	Satisfied (>3)	7	70
Total		10	100
No card withdrawals	Need improvement (≤3)	7	70
	Satisfied (>3)	3	30
Total		10	100
Foreign currency cash reservation	Need improvement (≤3)	9	90
	Satisfied (>3)	1	10
Total		10	100
Cross-border payments	Need improvement (≤3)	9	90
	Satisfied (>3)	1	10
Total		10	100
Network reservation	Need improvement (≤3)	6	60
	Satisfied (>3)	4	40
Total		10	100

Table 10: Summary of the investment/finance management.

Ease of investment/finance management			
Category	Ranking	Number	%
Investment/finance management services	Yes (1)	10	100
	No (0)	0	0
Total		10	100
Extensive m-banking products	Need improvement (≤3)	1	10
	Satisfied (>3)	9	90
Total		10	100
Detailed introduction	Need improvement (≤3)	2	20
	Satisfied (>3)	8	80
Total		10	100

Table 11: Summary of Livelihood payment.

Ease of use			
Category	Ranking	Number	%
Payment recharge (life payment, credit)	Need improvement (≤3)	1	10
	Satisfied (>3)	9	90
Total		10	100
Food and leisure (ticket, movie tickets)	Need improvement (≤3)	4	40
	Satisfied (>3)	6	60
Total		10	100
Traffic travel (paid fines, ticket purchases)	Need improvement (≤3)	3	30
	Satisfied (>3)	7	70
Total		10	100
Medical registration	(0, 10]	7	70
	(10, 20]	0	0
	(20, 30]	3	30
Total		10	100
Health management (health assessment, daily tasks)	Need improvement (≤3)	9	90
	Satisfied (>3)	1	10
Total		10	100

downloads reaching 320 million times. 50% of MB should improve their user evaluation, as shown in Table 12. According to the survey, user evaluation is the key factor for customers to download MB. Therefore, the bank should pay close attention to the contents of MB in order to improve users' evaluation.

The MB customer service is a very important factor to impact load. As shown in Table 10, all the observed banks have offered some consulting ways. In various ways of user feedback, such as message board, contact phone, online solution banks, and e-mail, can solve customers' needs more quickly. 10% of the banks have five consulting methods; 10%

Table 12: Summary of the popularity of the APP.

Website popularity			
Category	Ranking	Number	%
APP downloads	(0, 5000]	5	50
	(5000, 10000]	3	30
	(10000, 15000]	2	20
Total		10	100
User reviews	Need improvement (≤3)	5	50
	Satisfied (>3)	5	50
Total		10	100

Table 13: Summary of the customer service.

Ease of available common services			
Category	Ranking	Number	%
Consultation (message board, contact phone number, online answer, e-mail, etc.)	Method (=5)	1	10
	Method (=4)	0	0
	Method (=3)	1	10
	Method (=2)	4	40
	Method (=1)	4	40
Total		10	100
Feedback speed	Need improvement (≤3)	4	40
	Satisfied (>3)	6	60
Total		10	100
Feedback accuracy	Need improvement (≤3)	7	70
	Satisfied (>3)	3	30
Total		10	100

of the selected have three types of consulting; and most banks have one or two types of consulting, as shown in Table 13. According to the actual survey, 50% of MB need to speed up their feedback to improve customer satisfaction. In addition, 70% of banks need to improve the accuracy of their feedback; although most banks can answer it quickly, the feedback is irrelevant to the client's problem. Therefore, these are urgent problems for banks to settle which can improve the customers' satisfaction.

5. Solutions from DEMATEL-ANP Analysis

To further determine the factors which influence the adoption rate of MB, the DEMATEL-ANP is employed in this research. This approach relies on a team of experts to measure data. In the study, three experts were selected to discuss the motivations, decisions, and considerations that influence the adoption rate of MB. The three experts selected in this case are the company manager, technical engineer, and sales manager. The questionnaire is mainly used to evaluate the importance of the factors, and in the decision-making process, the evaluation criteria are sorted to obtain the direct relation matrix.

Step 1: the influencing factors, decision goals, standards, subcriteria, and alternatives of decision-making are shown in Table 14. In summary, nine criteria and 34 subcriteria are identified.

Step 2: in the process of choosing whether to download the MB, customers will be compared according to the 4-leveled scale of DEMATEL.

Step 3: internal dependencies are found in each index of the direct relationship matrix, as shown in Table 15. The columns sum {0, 3.67, 5.34, 8.33, 8, 3.68, 8, 26.66, 9.66} are derived from the above direct relationship matrix. Also, we can have the rows sum {11.34, 13.67, 12, 6.33, 6, 5, 5, 5.67, 8.33}. From this, we can get a maximum of the popularity of the columns 26.66. Then, according to formulas (1) and (2), the direct relation matrix is standardized, and the normalized direct relation matrix of Table 16 is obtained.

Step 4: finally, through formula (3), and using MATLAB tools to carry out operation, the comprehensive influence matrix is given in Table 17.

Step 5: with the help of formulas (4)–(6), computing center factors and reason factors are shown in Table 18.

Step 6: based on the previous steps, the matrix for influencing aspects is concluded by using the dataset $(D_i + R_i)$, $(D_i - R_i)$. The causal diagram for the total influencing factors is given in Figure 2. Due to the limited space, only five indicators are taken as examples. Here, the result indicates that A, B, C, and F are dispatchers and D, E, G, H, and I are receivers. Therefore, the visual effect, ease of use, security, and investment management index will have a great influence on the common services, convenience services, livelihood payment, popularity of the MB, and customer service index. From Table 14, it is seen that the visual effect (A) has the value of $(D_i - R_i = 0.583)$ and is considered to be an important factor affecting all the other factors with a high importance $(D_i + R_i = 0.583)$.

Table 14: The measures of the online services of mobile banking.

Goal	Aspects	Criteria
Online services of e-banking	A: visual effect	A_1 color assortment
		A_2 site aesthetic
		A_3 visual attraction
	B: ease of use	B_1 file size
		B_2 number of APPs to download
		B_3 finding convenience
		B_4 version (IOS, Android)
		B_5 problems with replacing the page
	C: security	C_1 online payment or remittance security method
		C_2 client confidentiality
		C_3 ensure safe innovation (fingerprint, face)
	D: common services	D_1 account management (balances, details)
		D_2 transfer money
		D_3 apply for a credit card
		D_4 deposit and loan management
		D_5 foreign currency management
	E: convenience services	E_1 mobile payment (collection, payment)
		E_2 no card withdrawals
		E_3 foreign currency cash reservation
		E_4 cross-border payments
		E_5 network reservation
	F: investment/finance management	F_1 investment/finance management services
		F_2 extensive m-banking products
		F_3 detailed introduction
	G: livelihood payment	G_1 payment recharge (life payment, credit)
		G_2 food and leisure (ticket, movie tickets)
		G_3 traffic travel (paid fines, ticket purchases)
		G_4 medical registration
		G_5 health management
	H: popularity of the APP	H_1 APP downloads
		H_2 user reviews
	I: customer service	I_1 consultation (message board, contact phone number, online answer, e-mail, etc.)
		I_2 feedback speed
		I_3 feedback accuracy

Table 15: The initial direct relation matrix for economic aspects.

	A	B	C	D	E	F	G	H	I
A	—	1.67	0	1.67	1.33	0.67	1.67	2.33	2
B	0	—	1	2	2	0.67	3	3.67	1.33
C	0	2	—	2	1.67	1	1.33	3.33	0.67
D	0	0	0.67	—	0.67	0	0.33	3.33	1.33
E	0	0	0.67	0.33	—	0	0	3.67	1.33
F	0	0	0.67	0	0	—	0	3.33	1
G	0	0	0.33	0	0	0	—	3.67	1
H	0	0	1	1	1.33	0.67	0.67	—	1
I	0	0	1	1.33	1	0.67	1	3.33	—

Table 16: The normalized direct relation matrix.

	A	B	C	D	E	F	G	H	I
A	0.000	0.063	0.000	0.063	0.050	0.025	0.063	0.087	0.075
B	0.000	0.000	0.038	0.075	0.075	0.025	0.113	0.138	0.050
C	0.000	0.075	0.000	0.075	0.063	0.038	0.050	0.125	0.025
D	0.000	0.000	0.025	0.000	0.025	0.000	0.012	0.125	0.050
E	0.000	0.000	0.025	0.012	0.000	0.000	0.000	0.138	0.050
F	0.000	0.000	0.025	0.000	0.000	0.000	0.000	0.125	0.038
G	0.000	0.000	0.012	0.000	0.000	0.000	0.000	0.138	0.038
H	0.000	0.000	0.038	0.038	0.050	0.025	0.025	0.000	0.038
I	0.000	0.000	0.038	0.050	0.038	0.025	0.038	0.125	0.000

Table 17: The comprehensive influence matrix.

	A	B	C	D	E	F	G	H	I
A	0.000	0.064	0.017	0.08	0.069	0.033	0.079	0.145	0.096
B	0.000	0.004	0.055	0.092	0.094	0.034	0.125	0.2	0.074
C	0.000	0.076	0.018	0.092	0.083	0.046	0.067	0.182	0.049
D	0.000	0.003	0.035	0.012	0.037	0.007	0.02	0.147	0.06
E	0.000	0.003	0.035	0.024	0.013	0.007	0.008	0.157	0.059
F	0.000	0.002	0.033	0.01	0.011	0.006	0.007	0.14	0.045
G	0.000	0.002	0.021	0.01	0.011	0.006	0.007	0.15	0.045
H	0.000	0.003	0.044	0.045	0.058	0.029	0.031	0.033	0.047
I	0.000	0.004	0.048	0.061	0.051	0.031	0.046	0.158	0.016

Table 18: The comprehensive influence matrix after influence.

	A	B	C	D	E	F	G	H	I	Di	Di + Ri	Di − Ri
A	0.000	0.064	0.017	0.08	0.069	0.033	0.079	0.145	0.096	0.583	0.583	0.583
B	0.000	0.004	0.055	0.092	0.094	0.034	0.125	0.2	0.074	0.678	0.839	0.517
C	0.000	0.076	0.018	0.092	0.083	0.046	0.067	0.182	0.049	0.613	0.919	0.307
D	0.000	0.003	0.035	0.012	0.037	0.007	0.02	0.147	0.06	0.321	0.747	−0.105
E	0.000	0.003	0.035	0.024	0.013	0.007	0.008	0.157	0.059	0.306	0.733	−0.121
F	0.000	0.002	0.033	0.01	0.011	0.006	0.007	0.14	0.045	0.254	0.453	0.055
G	0.000	0.002	0.021	0.01	0.011	0.006	0.007	0.15	0.045	0.252	0.642	−0.138
H	0.000	0.003	0.044	0.045	0.058	0.029	0.031	0.033	0.047	0.29	1.359	−1.022
I	0.000	0.004	0.048	0.061	0.051	0.031	0.046	0.158	0.016	0.415	0.906	−0.076
Ri	0.000	0.161	0.306	0.426	0.427	0.199	0.39	1.312	0.491			

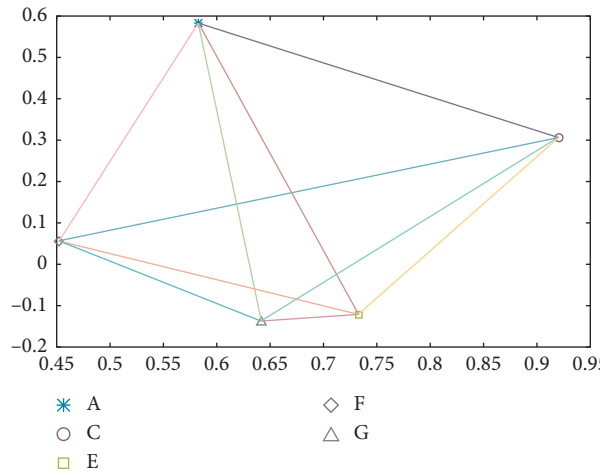


Figure 2: The causal diagram for the total influencing factors.

Table 19: The comprehensive influence matrix.

	A	B	C	D	E	F	G	H	I
A	0.000	0.064	0.017	0.08	0.069	0.033	0.079	0.145	0.096
B	0.000	0.004	0.055	0.092	0.094	0.034	0.125	0.2	0.074
C	0.000	0.076	0.018	0.092	0.083	0.046	0.067	0.182	0.049
D	0.000	0.000	0.035	0.012	0.037	0.007	0.000	0.147	0.06
E	0.000	0.000	0.035	0.024	0.013	0.007	0.008	0.157	0.059
F	0.000	0.000	0.033	0.000	0.011	0.006	0.007	0.14	0.045
G	0.000	0.000	0.021	0.000	0.011	0.006	0.007	0.15	0.045
H	0.000	0.000	0.044	0.045	0.058	0.029	0.031	0.033	0.047
I	0.000	0.004	0.048	0.061	0.051	0.031	0.046	0.158	0.016

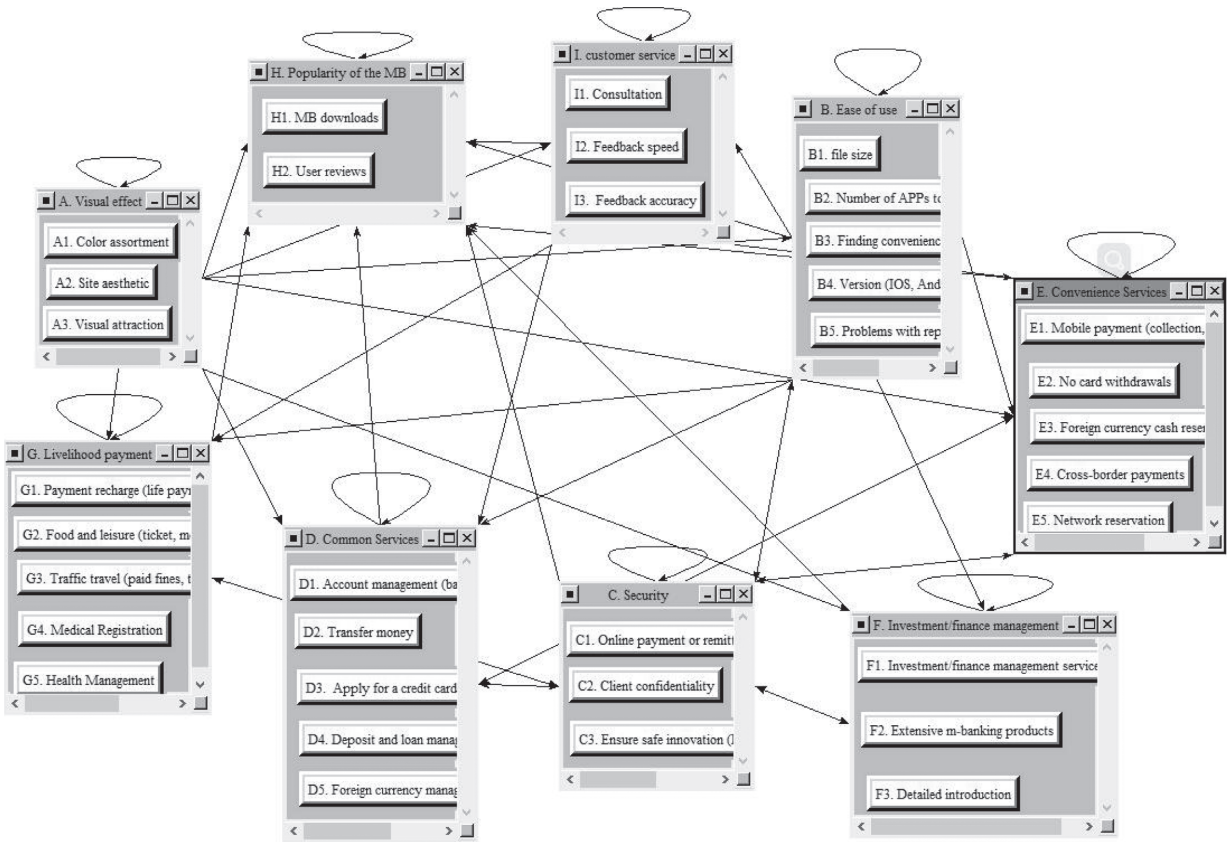


Figure 3: Network structure of the evaluation model.

Table 20: The pairwise comparison matrix.

Super decisions main window: influence factor.sdmod; cluster matrix view								
Cluster node labels	A: visual effect	B: ease of use	C: security	D: common services	E: convenience services	F: investment/ finance management	G: livelihood payment	H: popularity of the MB
A: visual effect	0.3333	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
B: ease of use	0.2909	0.2364	0.2527	0.0000	0.0000	0.0000	0.0000	0.0000
C: security	0.0000	0.4757	0.4898	0.5579	0.5579	0.4736	0.7085	0.0000
D: common services	0.1285	0.1059	0.1022	0.2772	0.2772	0.0000	0.0000	0.0000
E: convenience services	0.0645	0.0537	0.0396	0.1293	0.1293	0.0000	0.0000	0.0000
F: investment/ finance management	0.0542	0.0376	0.0422	0.0000	0.0000	0.4736	0.0000	0.0000
G: livelihood payment	0.0535	0.0458	0.0503	0.0000	0.0000	0.0000	0.2311	0.0000
G: popularity of the MB	0.0148	0.0201	0.0228	0.0354	0.0354	0.0526	0.0603	1.0000

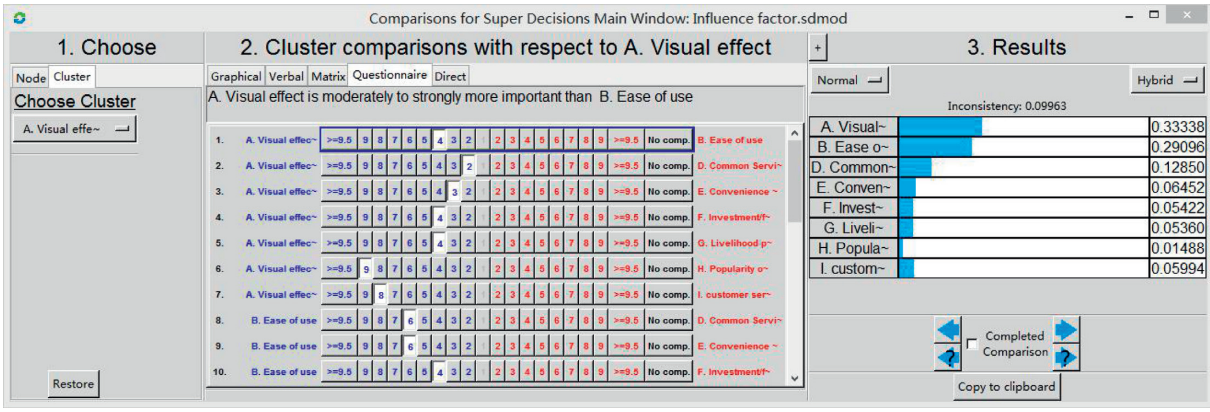
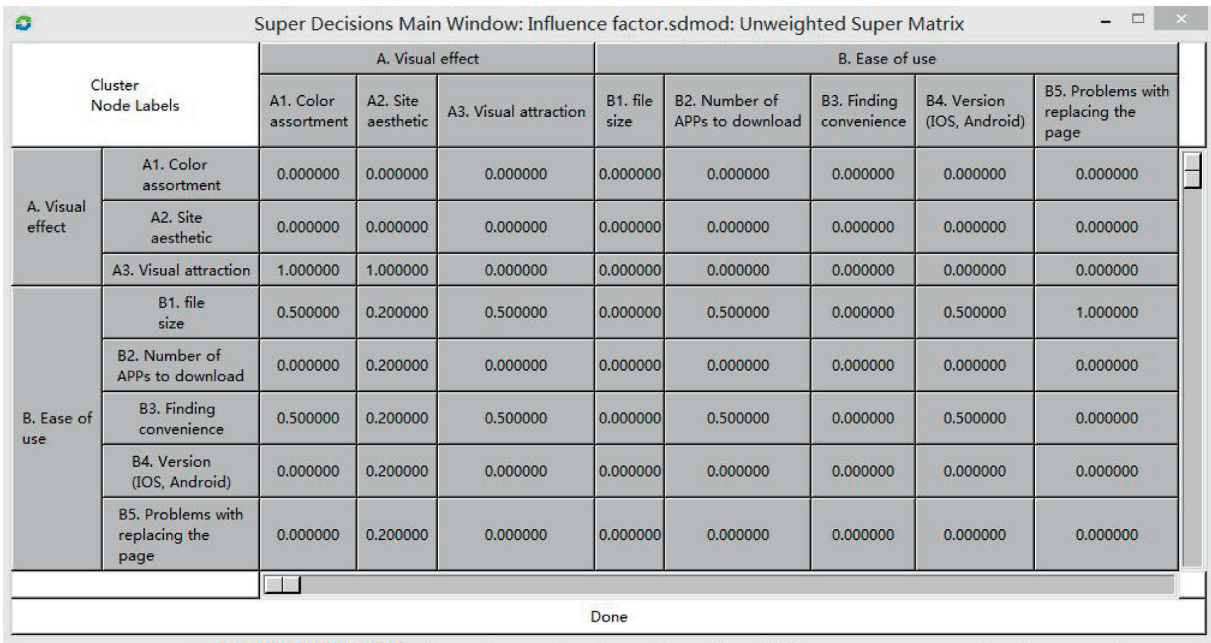


Figure 4: Display of the normalized weight vector and consistency index.



Cluster Node Labels		A1. Color assortment	A2. Site aesthetic	A3. Visual attraction	B1. file size	B2. Number of APPs to download	B3. Finding convenience	B4. Version (IOS, Android)	B5. Problems with replacing the page
A. Visual effect	A1. Color assortment	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
	A2. Site aesthetic	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
	A3. Visual attraction	1.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
B. Ease of use	B1. file size	0.500000	0.200000	0.500000	0.000000	0.500000	0.000000	0.500000	1.000000
	B2. Number of APPs to download	0.000000	0.200000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
	B3. Finding convenience	0.500000	0.200000	0.500000	0.000000	0.500000	0.000000	0.500000	0.000000
	B4. Version (IOS, Android)	0.000000	0.200000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
	B5. Problems with replacing the page	0.000000	0.200000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

Figure 5: Implementation of the unweighted super matrix.

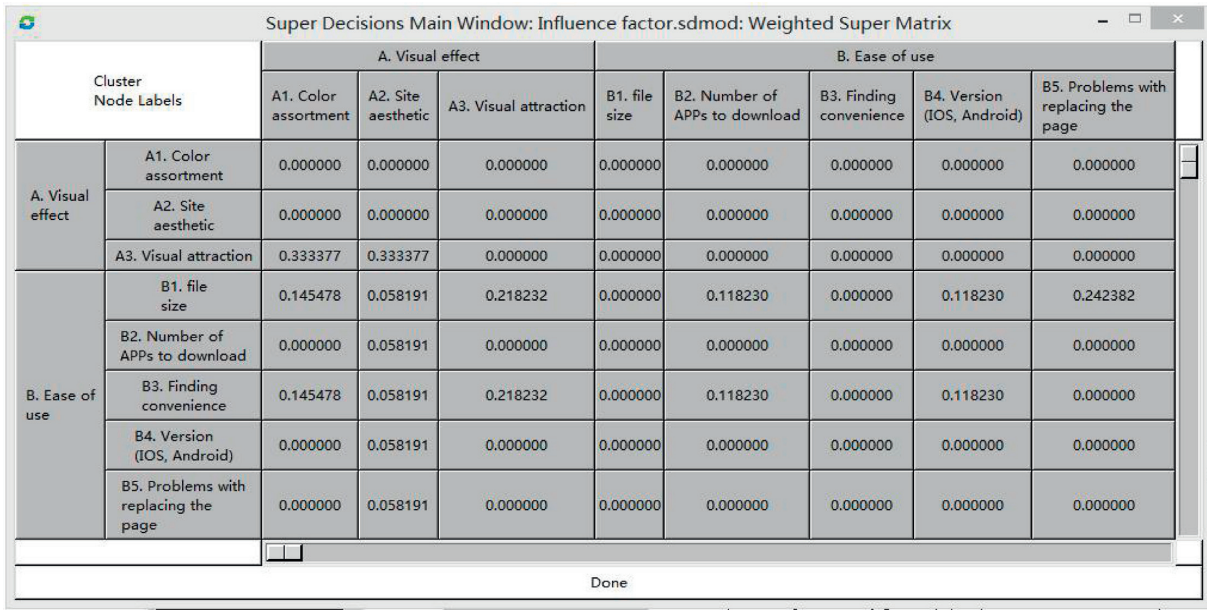
The ease of use (B) has $(D_i - R_i = 0.517)$ in the cause group, and its $(D_i + R_i = 0.839)$ is the second important indicator.

6. Conclusions and Suggestions for Future Research

This paper focuses on the factors influencing the adoption rate of MB, including the analysis of competition between five state-owned banks and five commercial banks in China's financial market. The purpose of this study is to (1) investigate the effectiveness of 10 banks in MB services; (2) compare the differences between state-owned banks and commercial banks in various indicators; and (3) provide relevant correction suggestions to relevant banks.

Step 7: according to the above results, it can be seen that there is a strong internal correlation among selected indicators. The internal dependence of other indicators is studied by the same research method. From the expert opinion and repeated tests, the threshold value = 0.03, remove the irrelevant association relationship and obtain the relationship shown in Table 19.

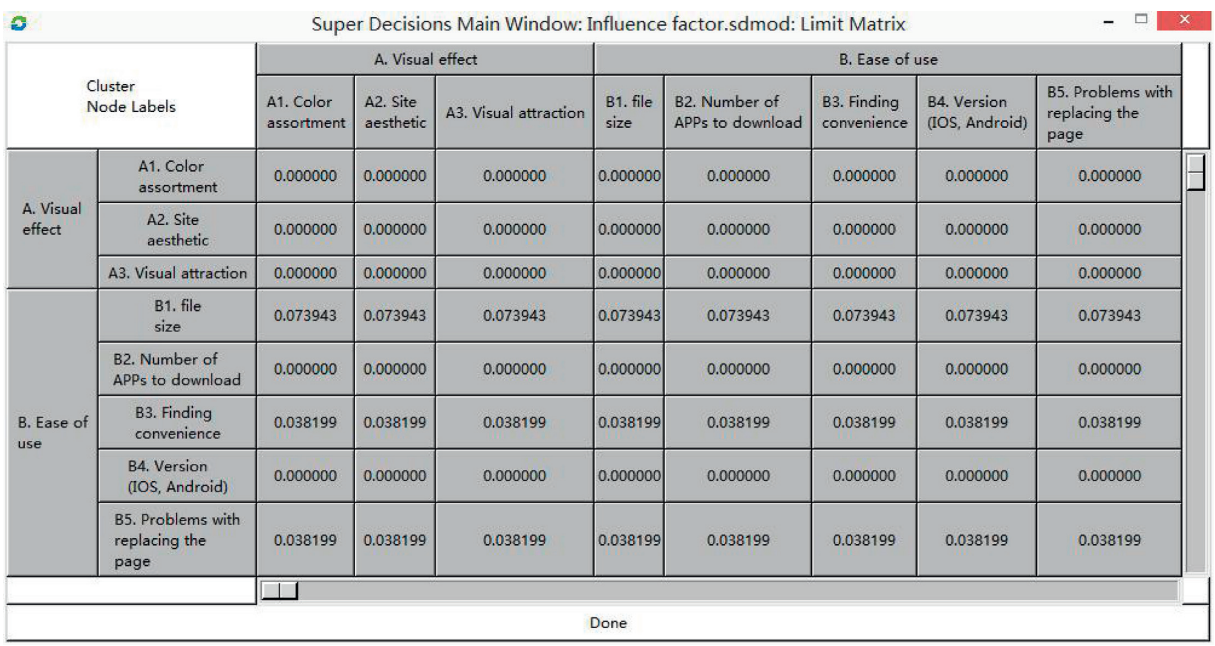
Step 8: relevant data generated by the DEMATEL method above is input into the ANP application software super decision, and the network structure of the evaluation model can be established through the external independence between factors, as shown in Figure 3. The application of the ANP method can better reflect the proportion of each index, while DM only shows the causal relationship between each factor and can give the judgment of the numerical


Super Decisions Main Window: Influence factor.sdmod: Weighted Super Matrix

Cluster Node Labels		A. Visual effect			B. Ease of use				
		A1. Color assortment	A2. Site aesthetic	A3. Visual attraction	B1. file size	B2. Number of APPs to download	B3. Finding convenience	B4. Version (IOS, Android)	B5. Problems with replacing the page
A. Visual effect	A1. Color assortment	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
	A2. Site aesthetic	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
	A3. Visual attraction	0.333377	0.333377	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
B. Ease of use	B1. file size	0.145478	0.058191	0.218232	0.000000	0.118230	0.000000	0.118230	0.242382
	B2. Number of APPs to download	0.000000	0.058191	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
	B3. Finding convenience	0.145478	0.058191	0.218232	0.000000	0.118230	0.000000	0.118230	0.000000
	B4. Version (IOS, Android)	0.000000	0.058191	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
	B5. Problems with replacing the page	0.000000	0.058191	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

Done

Figure 6: Implementation of the weighted supermatrix.


Super Decisions Main Window: Influence factor.sdmod: Limit Matrix

Cluster Node Labels		A. Visual effect			B. Ease of use				
		A1. Color assortment	A2. Site aesthetic	A3. Visual attraction	B1. file size	B2. Number of APPs to download	B3. Finding convenience	B4. Version (IOS, Android)	B5. Problems with replacing the page
A. Visual effect	A1. Color assortment	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
	A2. Site aesthetic	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
	A3. Visual attraction	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
B. Ease of use	B1. file size	0.073943	0.073943	0.073943	0.073943	0.073943	0.073943	0.073943	0.073943
	B2. Number of APPs to download	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
	B3. Finding convenience	0.038199	0.038199	0.038199	0.038199	0.038199	0.038199	0.038199	0.038199
	B4. Version (IOS, Android)	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
	B5. Problems with replacing the page	0.038199	0.038199	0.038199	0.038199	0.038199	0.038199	0.038199	0.038199

Done

Figure 7: Implementation of the limit matrix.

model. At the same time, using experts from different fields to grade evaluation indicators will greatly help the impartiality of the decision-making process. The next step is to use the ANP method to calculate the weight of each index.

Step 9: according to the comprehensive influence matrix, input the 10 mutual influences of the relationship between the indexes; due to space limitations, only capture some indexes between the pairwise comparison matrices, as shown in Table 20.

Step 10: through equations (7) and (8), the relative weight of each indicator can be obtained. Using Super Decisions software, the weight value of the above pairwise comparison matrix was input in the "Show New Priorities under the "Computation menu bar," and the normalized weight vector and consistency index value as shown in Figure 4 could be obtained. The consistency index value of the judgment matrix is 0.09963< 0.1, which meets the requirements and the weight is within the acceptable range.

Table 21: Implementation of the stable weight vector.

Aspects	Criteria	Local weight	Global weight
A: visual effect	A_1 color assortment	0.40124	0.216694
	A_2 site aesthetic	0.27514	0.148590
	A_3 visual attraction	0.32362	0.174771
B: ease of use	B_1 file size	0.49184	0.073943
	B_2 number of APPs to download	0.47049	0.028617
	B_3 finding convenience	0.25408	0.038199
	B_4 version (IOS, Android)	0.32988	0.014977
	B_5 problems with replacing the page	0.25408	0.381990
C: security	C_1 online payment or remittance security method	0.10124	0.000694
	C_2 client confidentiality	0.27234	0.148059
	C_3 ensure safe innovation (fingerprint, face)	0.12362	0.174771
D: common services	D_1 account management (balances, details)	0.34686	0.039817
	D_2 transfer money	0.19058	0.021877
	D_3 apply for a credit card	0.12685	0.014562
	D_4 deposit and loan management	0.16382	0.018806
	D_5 foreign currency management	0.17189	0.019732
E: convenience services	E_1 mobile payment (collection, payment)	0.29175	0.012195
	E_2 no card withdrawals	0.19667	0.008221
	E_3 foreign currency cash reservation	0.14828	0.006198
	E_4 cross-border payments	0.23017	0.009621
	E_5 network reservation	0.13313	0.005565
F: investment/finance management	F_1 investment/finance management services	0.37253	0.022659
	F_2 extensive m-banking products	0.15698	0.009548
	F_3 detailed introduction	0.47049	0.028617
G: livelihood payment	G_1 payment recharge (life payment, credit)	0.21616	0.009491
	G_2 food and leisure (ticket, movie tickets)	0.18494	0.008120
	G_3 traffic travel (paid fines, ticket purchases)	0.18494	0.008120
	G_4 medical registration	0.18494	0.008120
	G_5 health management	0.22903	0.010056
H: popularity of the MB	H_1 MB downloads	0.67012	0.030424
	H_2 user reviews	0.32988	0.014977
I: customer service	I_1 consultation (message board, contact phone number, online answer, e-mail, etc.)	0.91307	0.002626
	I_2 feedback speed	0.04346	0.000125
	I_3 feedback accuracy	0.04346	0.000125

Step 11: according to equations (9) and (10), all element level judgment matrices are entered in accordance with the above steps. If $C \cdot R \leq 0.10$, the consistency test is satisfied and the weight is within an acceptable range. If$C \cdot R \leq 0.10$, the result in this case is valid.

Step 12: the consistency of the matrix is tested. The Super Decision software, after passing through the consistency check, through the "Computation of main interface" menu bar under the "Unweighted Super Mater" and "Weighted Super Mater" bar menu command, can get a weighted matrix and weighted hypermatrix, as shown in Figures 5 and 6, respectively; The limit matrix can be obtained through the "Limit Matrix" menu command under the "Computation" menu bar in the main interface, as shown in Figure 7.

Step 13: through the above results, the final ranking of each index weight value can be obtained, as shown in Table 21.

The main data of this study were obtained through the comprehensive evaluation of MB services, which mainly includes the visual effect, ease of use, security, common service, convenience service, investment/finance management services, livelihood payment, the popularity of the MB, and customer service of the 10 banks which use DEMATEL-ANP evaluation on the MB services. According to the global weight obtained in Table 20, the four indicators for evaluating MB in online services are (1) visual effects, whose weight = (0.22, 0.15, 0.17), (2) ease of use, weight = (0.07, 0.03, 0.04, 0.01, 0.38), (3) security, weight = (0.00, 0.15, 0.17), and (4) common services, weight = (0.04, 0.02, 0.01, 0.02, 0.02). As shown in Table 1, the three factors that foreign scholars pay most attention to in the research on the adoption rate of MB are trust, risk perception, and behavioral intention. As a result, a number of important management influences are obtained: (1) MB in finance and investment service should enrich its product introduction and the related kinds of choices simplified the process of online purchase financial products at the same time. (2) Research shows that when users download MB, they will pay attention to other users' comments and the number of downloads of their software. (3) Banks should also enhance their services on people's livelihood payment so that users can pay their living expenses through MB and other related issues.

By comparing the data of state-owned banks and commercial banks, it can be concluded that the customer satisfaction of commercial banks is higher than that of state-owned banks. Commercial banks score higher on the ease of use, security, and common services than state-owned banks, for example, the file size in the ease of use, the software of state-owned banks size with an average of 53.86 M, and the commercial banks is 36.12. In terms of ensuring safe innovation in security, three of the five commercial banks have introduced fingerprint and facial recognition functions, which can ensure the security of user accounts and the convenience of operation. In the deposits and loans management in common services, the satisfaction rate of commercial banks is 60%, while that of the state-owned banks is only 20%. But in convenience services, significantly higher than the score of state-owned banks to commercial banks, in mobile payment, no card withdrawals, and network reservation of the three functions, the customer satisfaction of state-owned banks were higher than 50%, while in commercial banks, customers' satisfaction was only 40%, 20%, and 0%; in other words, the five commercial banks in the study does not exist in the function of network reservation. In livelihood payment, the popularity of MB, and customer service, three aspects of state-owned banks and commercial banks have their own advantages. For example, five state-owned banks can be fulfilled in this software for traffic travel (paid fines and ticket purchases), only two banks of commercial banks in providing the service. In terms of medical registration, for example, in Beijing, only 27 hospitals can be reserved by state-owned banks, while 57 hospitals can be reserved by commercial banks, more than double that of state-owned banks. Among the popularity of MB, the download volume of state-owned banks is much higher than that of commercial banks, and the user reviews of commercial banks is slightly higher than that of state-owned banks. Customer service in the kinds of consultation and feedback accuracy of the customer satisfaction of commercial banks were significantly higher than that of state-owned banks, and the feedback speed of customer satisfaction in state-owned banks is 80% and the commercial bank is only 20%.

Based on the results of this study, from the perspective of MB, there is still much room for banks to increase their MB adoption rate. We suggest that MB services of banks should be strengthened as follows: (1) banks should enhance their attention to users in terms of the softwires' color, structure, and visual appeal of MB; (2) to ensure that customer's personal information will not leak when participating in online banking, as well as to add the fingerprint and face unlock function, MB in the login, and payment security to further strengthen; (3) the size of the MB users to download the MB will focus on software and looking for convenience; at the same time, the number needed to download the MB software can significantly affect the downloads; banks should make the function of the MB concentration in a software; (4) generally, the management of deposit, loan, and foreign currency should be strengthened in the common services of MB, and the functions of credit card application, deposit and loan management, and foreign currency management should be increased; (5) enhancing the service quality and service speed of online customers service, reducing the waiting time of users, and improving customer satisfaction.

Data Availability

The main data of this study were obtained through the observation sheet of mobile banking services from the APP of the following banks: Industrial and Commercial Bank of China, Construction Bank of China, Bank of China, Agricultural Bank of China, Bank of Communications, China Postal Savings Bank, China Everbright Bank, China Merchants Bank, China Minsheng Bank, and Bank of Beijing.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The work was supported by the National Natural Science Foundation of China (nos. 71602008 and 71802021), Beijing Natural Science Foundation (no. 9184023), and Beijing Social Science Fund Research Project (nos. 16JDGLC032, 17JDGLB011, and 18GLB022).

References

[1] A. Shankar, C. Jebarajakirth, and M. Ashaduzzaman, "How do electronic word of mouth practices contribute to mobile banking adoption?" *Journal of Internet Banking and Commerce*, vol. 52, no. 1, pp. 1019-1020, 2020.

[2] T. Laukkanen and P. Cruz, *Comparing Consumer Resistance to Mobile Banking in Finland and Portugal. E-Business and Telecommunications*, Springer Berlin Heidelberg, Berlin, Germany, 2009.

[3] G. Mols, I. Brandes, V. Kessler et al., "Volume-dependent compliance in ards: proposal of a new diagnostic concept," *Intensive Care Medicine*, vol. 25, no. 10, pp. 1084–1091, 1999.

[4] L. Guttmann and W. Yan, "A research on audio-visual product industry ideology's security," *IEEE*, vol. 25, 2017.

[5] P. Luarn and H.-H. Lin, "Toward an understanding of the behavioral intention to use mobile banking," *Computers in Human Behavior*, vol. 21, no. 6, pp. 873–891, 2005.

[6] P. Cruz, L. Barretto Filgueiras Neto, P. Muñoz-Gallego, and T. Laukkanen, "Mobile banking rollout in emerging markets: evidence from Brazil," *International Journal of Bank Marketing*, vol. 28, no. 5, pp. 342–371, 2010.

[7] M. Polasik, "The use of electronic distribution channels of banking services in Poland," *Copernican Journal of Finance & Accounting*, vol. 2, pp. 139–152, 2013.

[8] P. Duarte, S. Costa e Silva, and M. B. Ferreira, "How convenient is it? Delivering online shopping convenience to enhance customer satisfaction and encourage e-WOM," *Journal of Retailing and Consumer Services*, vol. 44, pp. 161–169, 2018.

[9] A. A. Alalwan, Y. K. Dwivedi, N. P. P. Rana, and M. D. Williams, "Consumer adoption of mobile banking in Jordan," *Journal of Enterprise Information Management*, vol. 29, no. 1, pp. 118–139, 2016.

[10] B. Friedman, P. H. Khan, and D. C. Howe, "Trust online," *Communications of the Acm*, vol. 43, no. 12, pp. 34–40, 2000.

[11] T. Jacobs, *Multi-channel Banking: Banks Go Mobile*, Provenir Inc., London, UK, 2007.
[12] M. S. Featherman and P. A. Pavlou, "Predicting e-services adoption: a perceived risk facets perspective," *International Journal of Human-Computer Studies*, vol. 59, no. 4, pp. 451–474, 2003.
[13] D. J. Kim, D. L. Ferrin, and H. R. Rao, "A trust-based consumer decision-making model in electronic commerce: the role of trust, perceived risk, and their antecedents," *Decision Support Systems*, vol. 44, no. 2, pp. 544–564, 2008.
[14] K. S. Lee, H. S. Lee, and S. Y. Kim, "Factors influencing the adoption behavior of mobile banking: a South Korean perspective," *Journal of Internet Banking and Commerce*vol. 12, no. 2, 2007.
[15] P. Hanafizadeh, M. Behboudi, and A. Abedini Koshksaray, "Mobile-banking adoption by iranian bank clients," *Telematics and Informatics*, vol. 31, no. 1, pp. 62–78, 2014.
[16] A. Jalilvand Shirkhani Tabar and B. Datta, "Factors affecting mobile payment adoption intention: an Indian perspective," *Global Top Performers*, vol. 19, no. 3, pp. 72–89, 2018.
[17] M. A. Shareef, A. Baabdullah, S. Dutta, V. Kumar, and Y. K. Dwivedi, "Consumer adoption of mobile banking services: an empirical examination of factors according to adoption stages," *Journal of Retailing and Consumer Services*, vol. 43, pp. 54–67, 2018.
[18] A. Shankar, B. Datta, and C. Jebarajakirthy, "Are the generic scales enough to measure service quality of mobile banking? a comparative analysis of generic service quality measurement scales to mobile banking context," *Services Marketing Quarterly*, vol. 40, no. 3, pp. 224–244, 2019.
[19] C. Changchit, R. Lonkani, and J. Sampet, "Mobile banking: exploring determinants of its adoption," *Journal of Organizational Computing & Electronic Commerce*, vol. 27, no. 1, 2017.
[20] X. Luo, H. Li, J. Zhang, and J. P. Shim, "Examining multi-dimensional trust and multi-faceted risk in initial acceptance of emerging technologies: an empirical study of mobile banking services," *Decision Support Systems*, vol. 49, no. 2, pp. 222–234, 2010.
[21] M. S. Y. Lee, P. J. Mcgoldrick, K. A. Keeling, and J. Doherty, "Using zmet to explore barriers to the adoption of 3g mobile banking services," *International Journal of Retail & Distribution Management*, vol. 31, no. 6, pp. 340–348, 2003.
[22] W. W. Chin, "Issues and opinions on structural equation modeling," *Embo Journal*, vol. 11, no. 12, pp. 4261–4272, 2003.
[23] D. Sahoo and S. S. Pillai, "Role of mobile banking servicescape on customer attitude and engagement: an empirical investigation in India," *International Journal of Bank Marketing*, vol. 28, no. 4, 2017.
[24] T. Kuisma, T. Laukkanen, and M. Hiltunen, "Mapping the reasons for resistance to internet banking: a means-end approach," *International Journal of Information Management*, vol. 27, no. 2, pp. 75–85, 2007.
[25] V. Bhatiasevi, "An extended utaut model to explain the adoption of mobile banking," *Information Development*, vol. 32, no. 4, 2016.
[26] H.-F. Lin, "Determining the relative importance of mobile banking quality factors," *Computer Standards & Interfaces*, vol. 35, no. 2, pp. 195–204, 2013.
[27] I. Ajzen and T. J. Madden, "Prediction of goal-directed behavior: attitudes, intentions, and perceived behavioral control," *Journal of Experimental Social Psychology*, vol. 22, no. 5, pp. 453–474, 1986.
[28] A. A. Alalwan, Y. K. Dwivedi, and N. P. Rana, "Factors influencing adoption of mobile banking by Jordanian bank customers: extending UTAUT2 with trust," *International Journal of Information Management*, vol. 37, no. 3, pp. 99–110, 2017.
[29] N. Sobti, "Impact of demonetization on diffusion of mobile payment service in India," *Journal of Advances in Management Research*, vol. 16, no. 4, pp. 472–497, 2019.
[30] H. E. Riquelme and R. E. Rios, "The moderating effect of gender in the adoption of mobile banking," *International Journal of Bank Marketing*, vol. 28, no. 5, pp. 328–341, 2010.
[31] A. A. Shaikh and H. Karjaluoto, "Mobile banking adoption: a literature review," *Telematics and Informatics*, vol. 32, no. 1, pp. 129–142, 2015.
[32] A. Shankar and P. Kumari, "Factors affecting mobile banking adoption behavior in India," *Journal of Internet Banking and Commerce*, vol. 21, no. 1, pp. 1–24, 2016.
[33] H.-F. Lin, "An empirical investigation of mobile banking adoption: the effect of innovation attributes and knowledge-based trust," *International Journal of Information Management*, vol. 31, no. 3, pp. 252–260, 2011.
[34] A. Giovanis, P. Athanasopoulou, C. Assimakopoulos, and C. Sarmaniotis, "Adoption of mobile banking services," *International Journal of Bank Marketing*, vol. 37, no. 5, pp. 1165–1189, 2019.
[35] T. Zhou, *Understanding Users' Initial Trust in Mobile Banking: An Elaboration Likelihood Perspective*, Elsevier Science Publishers, Berlin, Germany, 2012.
[36] G. Kim, B. Shin, and H. G. Lee, "Understanding dynamics between initial trust and usage intentions of mobile banking," *Information Systems Journal*, vol. 19, no. 3, pp. 283–311, 2009.
[37] C. Zhou, Z. Yang, and L. Jiang, "When do formal control and trust matter? a context-based analysis of the effects on marketing channel relationships in China," *Industrial Marketing Management*, vol. 40, no. 1, pp. 86–96, 2011.
[38] J.-C. Gu, S.-C. Lee, and Y.-H. Suh, "Determinants of behavioral intention to mobile banking," *Expert Systems with Applications*, vol. 36, no. 9, pp. 11605–11616, 2009.
[39] S. Brown and K. Taylor, "Early influences on saving behaviour: analysis of British panel data," *Journal of Banking & Finance*, vol. 62, pp. 1–14, 2016.
[40] G. Baptista and T. Oliveira, "Why so serious? gamification impact in the acceptance of mobile banking services," *Internet Research*, vol. 27, no. 1, pp. 118–139, 2017.
[41] A. A. Alalwan, N. P. Rana, Y. K. Dwivedi, B. Lal, and M. D. Williams, "Adoption of mobile banking in Jordan: exploring demographic differences on customers' perceptions," *Open and Big Data Management and Innovation*, vol. 9373, pp. 13–23, 2015.
[42] T. Zhou, Y. Lu, and B. Wang, "Integrating TTF and UTAUT to explain mobile banking user adoption," *Computers in Human Behavior*, vol. 26, no. 4, pp. 760–767, 2010.
[43] C. Martins, T. Oliveira, and A. Popovič, "Understanding the internet banking adoption: a unified theory of acceptance and use of technology and perceived risk application," *International Journal of Information Management*, vol. 34, no. 1, pp. 1–13, 2014.
[44] O. Bayazit and B. Karpak, "An analytical network process-based framework for successful total quality management (tqm): an assessment of Turkish manufacturing industry readiness," *International Journal of Production Economics*, vol. 105, no. 1, pp. 79–96, 2007.
[45] B. Karpak and I. Topcu, "Small medium manufacturing enterprises in Turkey: an analytic network process framework

for prioritizing factors affecting success," *International Journal of Production Economics*, vol. 125, no. 1, pp. 60–70, 2010.

[46] C.-Y. Huang, J. Z. Shyu, and G.-H. Tzeng, "Reconfiguring the innovation policy portfolios for taiwan's sip mall industry," *Technovation*, vol. 27, no. 12, pp. 744–765, 2007.

[47] S. Rahman and N. Subramanian, "Factors for implementing end-of-life computer recycling operations in reverse supply chains," *International Journal of Production Economics*, vol. 140, no. 1, pp. 239–248, 2012.

[48] C. Y. Huang, Y. W. Liu, and G. H. Tzeng, "Derivations of factors influencing the technology acceptance of smart TV by using the DEMATEL based network process," *Intelligent Decision Technologies*, vol. 140, 2012.

[49] J. L. Yang and G.-H. Tzeng, "An integrated mcdm technique combined with dematel for a novel cluster-weighted with anp method," *Expert Systems with Applications*, vol. 38, no. 3, pp. 1417–1424, 2011.

[50] G. Büyüközkan and D. Öztürkcan, "An integrated analytic approach for six sigma project selection," *Expert Systems with Applications*, vol. 37, no. 8, pp. 5835–5847, 2010.

[51] T.-M. Yeh and Y.-L. Huang, "Factors in determining wind farm location: integrating GQM, fuzzy DEMATEL, and ANP," *Renewable Energy*, vol. 66, no. 3, pp. 159–169, 2014.

[52] Y.-H. Chang, W.-M. Wey, and H.-Y. Tseng, "Using anp priorities with goal programming for revitalization strategies in historic transport: a case study of the alishan forest railway," *Expert Systems with Applications*, vol. 36, no. 4, pp. 8682–8690, 2009.

[53] W.-W. Wu, "Choosing knowledge management strategies by using a combined ANP and DEMATEL approach," *Expert Systems with Applications*, vol. 35, no. 3, pp. 828–835, 2008.

[54] G. Büyüközkan and S. Güleryüz, *A New Integrated Intuitionistic Fuzzy Group Decision Making Approach for Product Development Partner Selection*, Pergamon Press, London, UK, 2016.

[55] G. Büyüközkan and G. Çifçi, "A novel hybrid mcdm approach based on fuzzy dematel, fuzzy anp and fuzzy topsis to evaluate green suppliers," *Expert Systems with Applications*, vol. 39, no. 3, pp. 3000–3011, 2012.

Hindawi
Discrete Dynamics in Nature and Society
Volume 2020, Article ID 5138135, 10 pages
https://doi.org/10.1155/2020/5138135

Research Article

Study on ETFEE in the BTH Region Based on the Window-SBM-Undesirable Model

Feng Ren, Xin Yu, and Quan Li

School of Business and Administration, North China Electric Power University, No. 689, Huadian Road, Lianchi District, Baoding 071003, Hebei, China

Correspondence should be addressed to Feng Ren; renfeng2002@126.com

Received 3 April 2020; Revised 29 July 2020; Accepted 4 August 2020; Published 6 October 2020

Guest Editor: Fuqiang Gu

Research on ecological total-factor energy efficiency (ETFEE) is conducive to energy conservation, emission reduction, and ecological protection. This paper focuses on the measurement and decomposition of the ETFEE in the Beijing-Tianjin-Hebei (BTH) region in China. In order to measure the ETFEE values, the window technology is combined with a nonradial and nonoriented SBM-undesirable model considering undesirable outputs to overcome the defect of insufficient data of research objects and ensure the calculation process to be implemented. The findings show that Beijing and Tianjin are DEA-efficient, while Hebei is not. The technological progress rates of Beijing and Tianjin reach up to 11.92% and 14.96%, while that of Hebei retrogresses by 4.47%. The scale efficiencies of Beijing, Tianjin, and Hebei are 97.75%, 86.60%, and 93.81%, respectively, which means that there are potentials for further optimization in the energy structures. The impulse response results between the energy structure and the ETFEE show that the proportions of coal and petroleum have negative effects on the ETFEE, while that of natural gas has a positive effect. The research results can provide reference for decision makers to formulate regional development plans.

1. Introduction

Beijing-Tianjin-Hebei (BTH) urban agglomeration is one of the top three world-class urban agglomerations in China, with a total area of 216,000 square kilometers (2.25% of the country's land area) and a population of about 110 million (8.1% of the country's total population), creating 9.46% of the gross domestic product (source: China Statistical Yearbook 2019), playing a very important role in China's overall development strategy. However, the economic and social development of BTH urban agglomeration has been achieved via large-scale yet inefficient consumption of energies. In 2014, China's total energy consumption reached 4.003 billion tons of standard coal, among which 442.966 million tons of standard coal were consumed in the BTH region, accounting for 11.07% of the country's total. China's coal consumption in 2015 was 2.762 billion tons of standard coal, and BTH's share accounted for 13.18%, among which the share of the Hebei province accounted for 81.4% of the total in the BTH region (source: China Energy Statistical Yearbook 2016). The superfluous inefficient consumption of energy has led to the deterioration of the ecological environment in the BTH region. In recent years, China's Air Quality Index (AQI) has frequently stayed beyond the scale, and 6 to 8 of the top 10 most polluted cities in China are located in the BTH region. China's national "13th Five-Year Plan," as well as the "13th Five-Year Plans" of Beijing, Tianjin, and Hebei Prov., clearly put forward the goals of promoting the transformation of energy production and utilization methods, optimizing the energy supply structure, and improving energy efficiency. Energy utilization should meet the needs of economic and social development without destructive impacts on the environment. So, how to improve the ecological total-factor energy efficiency (ETFEE) in this urban agglomeration has become an important issue, which should be considered seriously.

The importance of studying the ETFEE in BTH urban agglomeration under the background of supply-side structural reform can be represented in the following aspects: (1) be conducive to alleviating the pressure of

energy shortage. Reliable energy supply is a major strategic issue related to the sustainable development for any region. Increasing energy efficiency can help to reduce energy consumption, alleviate the pressure of energy shortage, and contribute to the realization of reliable energy supply in this region. (2) Be helpful to reduce carbon emissions. In 2017, the world's energy-related carbon emissions rebounded, having increased by 1.7% (the average growth rate in the past decade is only 1.3%). According to authoritative reports, China was the biggest contributor to the carbon emission increase [1]. China has been facing the growing international pressure on carbon emission reductions and has set the carbon reduction target for 2030. Carbon emission reduction will undoubtedly be a hard constraint condition on the economic and social development in China [2, 3]. Obviously, the increase of the ETFEE in the BTH region can reduce carbon emissions and loosen the constraint for the long-term development in this region. (3) Be beneficial to protect the ecological environment. Increasing energy efficiency and reducing energy consumption will help to protect the ecological environment in the region. (4) Be conducive to grasping the direction of the energy structural reform. We need a large amount of data and reliable conclusions as the basis of scientific decision-making for the structural reform of the energy supply side. The study on the ETFEE in BTH urban agglomeration can enrich the theory and practice on the energy structural reform. (5) Provide references for other urban agglomerations. China has many urban agglomerations, such as the Pearl River Delta urban agglomeration, Yangtze River Delta urban agglomeration, Chengdu and Chongqing urban agglomeration, and the Middle Reaches of the Yangtze River urban agglomeration. These urban agglomerations may also come up against the similar ETFEE problems during their developments. The research methods and findings of this paper can provide references for these urban agglomerations.

In summary, the study on the ETFEE in BTH urban agglomeration is of great practical and academic values for energy conservation, emission reduction, environmental protection, and even for sustainable development of economy and society in China.

2. Literature Review

At present, energy efficiency research methods can be divided into three categories, namely, single-factor energy efficiency method, index analysis method, and total-factor energy efficiency method.

Single-factor energy efficiency is defined as the amount of energy consumption per unit of GDP. Single-factor energy efficiency has some advantages, such as being easy to calculate and make comparison between different regions. However, single-factor energy efficiency also has congenital defects. For example, it is vulnerable to factors such as climate, energy price, and demographic structure [4–6]. In 1997, Nagata proposed the concept of "real energy intensity," which refers to the energy efficiency after removing the influences of the aforementioned factors [7]. Besides, the use of single-factor energy efficiency may lead to the loss of structural information.

Some index analysis methods for energy efficiency have been proposed and applied to energy efficiency research studies. Su et al. [8] investigated the drivers of carbon emission changes of Singapore through a structural decomposition analysis (SDA); Román and Colinet [9], Zhao et al. [10], and He et al. [11] analyzed the driving forces behind the energy consumption changes via SDA. Zha et al. [12] explored whether the changes in the technology gap are one of the key driving factors of CO_2 emissions by index decomposition analysis (IDA); Liu and Xu [13] used IDA methods to decompose energy intensity change in China's metallurgical industry, while Xie and Lin [14] studied the energy intensity changes in China's food industry. Rüstemoğlu and Andrés [15] explored the energy-related CO_2 emissions in Brazil and Russia, and the refined Laspeyres index method was applied, and both aggregated and sectoral changes in CO_2 emissions decomposed; Shang et al. [16] analyzed the changing industrial water use in Tianjin through a refined Laspeyres model. González et al. [17] and González [18] used Divisia decomposition to explore the evolution of real energy efficiency in the European Union and the influence that the changes in sectoral composition in most EU economie, respectively. Moreover, Choi and Oh [19] studied the energy efficiency of the manufacturing industry of Korea, and Liao and Wei [20] examined China's aggregate energy consumption by using the Divisia approach. Although there are many improvements in index analysis methods compared with the single-factor energy efficiency method, they still cannot reflect the influences of the substitution effect between various factors on energy efficiency.

In 2006, Hu and Wang [21] first put forward the concept of total-factor energy efficiency (TFEE) and used the data envelopment analysis (DEA) method to calculate TFEE within the framework of capital, labor, and energy. In essence, TFEE is the ratio of the target energy input to the actual energy input, so calculating the target energy input is the critical step for TFEE calculation. DEA method is a nonparametric frontier method suitable for calculating the target energy input. Although stochastic frontier analysis (SFA) can also be used to calculate TFEE, it is not often chosen because of its function set artificially; otherwise, it may increase the risk of utilizing SFA to some extent. In addition, SFA cannot solve the problem of multi-input and multioutput [22, 23]. Therefore, after Hu and Wang, DEA method has been utilized by more and more researchers and has been a mainstream method for TFEE calculation.

Most of the existing literature directly added energy as a production factor into the DEA model and took the calculated efficiency as TFEE without distinguishing the production factors of labor and capital [24, 25]. In this case, the so-called TFEE cannot highlight the characteristics of energy [26]. To solve this problem, Tone [27] proposed a nonradial and nonoriented slack-based measure (SBM) model in 2001. The SBM method can separate the energy efficiency from

total-factor productivity and obtain the true energy efficiency. Besides, SBM can also consider the input and output slack variables simultaneously. The traditional DEA model does not take undesirable outputs into account. Many studies assumed that outputs were all desirable ones, but it is not true. For example, in a thermal power plant, a large amount of undesirable outputs such as carbon dioxide, sulfur dioxide, and soot is being emitted, while coal is burned and converted to electricity. In 1983, Pittman first incorporated pollution factors into the efficiency model [28]. To sum up, we take the total-factor energy efficiency considering undesirable outputs as ecological total-factor energy efficiency (ETFEE). Since SBM can take energy saving and emission reduction into account simultaneously, we utilize this method in our research.

As for the TFEE in the BTH region in China, some researchers have studied it in recent years. For example, Wang et al. [29] used the DEA model to calculate the TFEE in the BTH region using data from the year 2005 to 2009 and found that technological progress was the leading force in the changing process of TFEE. Feng et al. [30] used the SBM model to measure the energy efficiency in the BTH region, analyzed the influencing factors of energy efficiency, and found that the most significant influencing factors were energy consumption structure, industrial structure, economic development level, and foreign investment. Wang [31] evaluated the energy efficiency in the BTH region by constructing a regional energy efficiency index system and concluded that the coordinated development of energy efficiencies among Beijing, Tianjin, and Hebei Prov. reached a coordinated development interval.

The existing literature on the TFEE in the BTH region has the following shortcomings: (1) the DEA method used in some studies failed to meet the empirical rule of thumb; (2) the effects of undesired outputs were not taken into account in some studies of energy efficiency; (3) almost none of the previous studies has comprehensively studied ETFEE in BTH urban agglomeration in the context of energy supply-side structural reform; and (4) most studies directly take energy structure as one of the factors affecting energy efficiency without considering the lagged effect and dynamic influence between the energy structure and the energy efficiency.

This study intends to make supplements to the existing literature in the following aspects: (1) construct an improved Window-SBM-undesirable model to meet the requirements for calculating the ETFEE in BTH urban agglomeration; (2) measure the recent ETFEE in BTH urban agglomeration scientifically from the perspectives of energy saving and emission reduction with undesirable outputs taken into consideration; (3) study the potential of energy conservation and emission reduction in the BTH urban agglomeration under the background of the energy supply-side structural reform; and (4) analyze the lag effect and dynamic influence relationship between the energy structure and the energy efficiency. The rest of this paper is organized as follows: firstly, the modeling process is given; then, the next part is empirical analysis, including data description and result analysis; and the last part is conclusions.

3. Modeling Process

3.1. Window Technology. DEA model should meet the basic rule of thumb: the number of decision-making units (DMUs) should not be less than the product of the number of input variables and that of output variables, and meanwhile, it should not be less than three times the sum of the number of input variables and that of output variables, too [32]. That is,

$$C \geq \max\{A \times B,\ 3 \times (A + B)\}, \tag{1}$$

where A and B represent the numbers of input variables and output variables, respectively, and C represents the number of DMUs.

However, the basic rule of thumb cannot be met since there are only three DMUs in this research. To deal with this problem effectively, we combine window technology with the SBM method to establish a Window-SBM model.

The Window-SBM method uses the data of each DMU repeatedly during different periods of efficiency measurement. Firstly, we select the window width d and then get $d \times J$ DMUs in a window (J is the number of original DMUs). If total length of time is T, then the number of DMUs will reach $d \times J \times (T - d + 1)$. For each DMU, we establish $T - d + 1$ windows for efficiency measurement. We will get d efficiency values for every DMU in the mth window ($m = 1, 2, \ldots, T - d + 1$). Starting from the time point $t = 1$, we can get d efficiency values in the first window, then move to the next time point $t = 2$, and so on, till $t = T - d + 1$. Finally, the average value of the efficiencies at each time point is taken as the final value of the evaluated DMU efficiency.

3.2. SBM-Undesirable Model. SBM-undesirable model deals with the situations when undesirable outputs exist, by setting the improvements of undesirable outputs in the opposite direction to the desirable outputs, which will result in more good outputs and less bad outputs. The SBM model with undesirable outputs is expressed as follows:

$$\min \rho = \frac{1 - (1/m)\sum_{i=1}^{m} s_i^- / x_{i0}}{1 + (1/q_1 + q_2)\left(\sum_{r=1}^{q_1} s_r^+ / y_{r0} + \sum_{t=1}^{q_2} s_t^{b-} / b_{t0}\right)},$$

$$\text{s.t.} \begin{cases} \sum_{j=1}^{n} x_{ij}\lambda_j + s_i^- = x_{i0}, & (i = 1, 2, \ldots, m), \\ \sum_{j=1}^{n} y_{rj}\lambda_j - s_r^+ = y_{r0}, & (r = 1, 2, \ldots, q_1), \\ \sum_{j=1}^{n} b_{tj}\lambda_j + s_t^{b-} = b_{t0}, & (t = 1, 2, \ldots, q_2), \\ \lambda_j, s_i^-, s_r^+, s_t^{b-} \geq 0, \end{cases} \tag{2}$$

where ρ is the efficiency value; n is the number of DMUs; m, q_1, and q_2 are the numbers of the inputs, the desirable outputs, and the undesirable outputs, respectively; s_i^-, s_r^+, and

s_t^{b-} are the slack variables of the inputs, the desirable outputs, and the undesirable outputs, respectively; λ_j are the intensive parameters; x_j, y_j, and b_j represent the variables of the inputs, the desirable outputs, and the undesirable outputs of DMU_j, respectively; and x_0, y_0, and b_0 represent the values of the inputs, the desirable outputs, and the undesirable outputs of DMU_0, respectively.

From model (2) above, we can see that the slack variables of the inputs and the outputs are put into the objective function of the SBM model, and the inefficiency value can be directly measured when compared with the optimal production frontier. It is in this way that the slack problem gets solved, and accordingly, the efficiency evaluation of the undesirable outputs can be effectively dealt with synchronously.

3.3. Super SBM-Undesirable Model. In order to solve the comparability problem among effective DMUs, this study combines the SBM-undesirable model with the super DEA method to construct a super SBM-undesirable model which is expressed as follows:

$$\min \rho = \frac{1 + (1/m)\sum_{i=1}^{m} s_i^- / x_{i0}}{1 - (1/q_1 + q_2)\left(\sum_{r=1}^{q_1} s_r^+ / y_{r0} + \sum_{t=1}^{q_2} s_t^{b-} / b_{t0}\right)},$$

$$\text{s.t.} \begin{cases} \sum_{j=1, j\neq j_0}^{n} x_{ij}\lambda_j - s_i^- \le x_{i0}, & (i = 1, 2, \ldots, m), \\ \sum_{j=1, j\neq j_0}^{n} y_{rj}\lambda_j + s_r^+ \ge y_{r0}, & (r = 1, 2, \ldots, q_1), \\ \sum_{j=1, j\neq j_0}^{n} b_{tj}\lambda_j + s_t^{b-} \le b_{t0}, & (t = 1, 2, \ldots, q_2), \\ 1 - \dfrac{1}{q_1 + q_2}\left(\sum_{r=1}^{q_1}\left(\dfrac{s_r^+}{y_{r0}}\right) + \sum_{t=1}^{q_2}\left(\dfrac{s_t^{b-}}{b_{t0}}\right)\right) > 0, \\ \lambda_j, s_i^-, s_r^+, s_t^{b-} \ge 0, & (j = 1, 2, \ldots, n; j \neq j_0). \end{cases} \tag{3}$$

The explanations of the parameters in model (3) can be referred to as in model (2). The effectiveness of the energy efficiency in model (3) is determined based on the following rules: (1) if $\rho < 1$, then DMU_0 is non-DEA-efficient; (2) if $\rho \ge 1$, then DMU_0 is DEA-efficient, and the bigger the value, the higher the efficiency.

4. Empirical Analysis

4.1. Data Description. The panel data of Beijing, Tianjin, and Hebei Prov. from 2001 to 2016 are selected in this study. The input data include capital, labor, and energy; the output data include GDP and sulfur dioxide emissions, of which sulfur dioxide is taken as an undesirable output (sources: China Statistical Yearbook and China Energy Statistical Yearbook). The relevant data are described as follows.

(1) Capital: generally speaking, studies involving capital investment should use capital stock as a proxy variable for capital input. However, capital stock data normally are not available in the official statistics released by the Chinese government; China's total capital stock could only be estimated in many literature studies. For example, Zhang et al. [33] put forward the perpetual inventory method to estimate capital stock. As in most literature studies, this study calculates the capital stock data of Beijing, Tianjin, and Hebei Prov. based on the perpetual inventory method proposed by Zhang et al. (2) Labor: the total number of employed people in Beijing, Tianjin, and Hebei Prov. from 2001 to 2016 is used as labor input data. The total number of employed people includes the numbers of people employed in state-owned enterprises, private enterprises, and self-employed businesses. (3) Energy: the total energy consumptions from 2001 to 2016 are used as energy input data. Different from the previous studies, the energy inputs in this study include both primary energy input and secondary energy input. The energies consumed include coal, coke, crude oil, gasoline, kerosene, diesel, fuel oil, natural gas, and electricity. The consumption data of all kinds of energies are converted into data expressed by standard coal before adding them up. (4) GDP: the GDP data from 2001 to 2016 are used as desirable output data. Based on the data of 2001, this study calculates GDP data of Beijing, Tianjin, and Hebei Prov. from 2001 to 2016 at constant prices. (5) Sulfur dioxide emissions: according to Fare et al. [34], sulfur dioxide is the most critical pollutant to affect outputs. Considering the availability of data, this study selects sulfur dioxide as a representative of undesirable outputs.

4.2. Spearman Correlation Test. The Spearman rank correlation coefficient is a distribution-independent rank statistical parameter to measure the strength of the relationship between two variables. Considering that the input and output variables should have a certain correlation, in order to ensure the authenticity of the efficiency measurement, the Spearman correlation test is carried out on the input-output variables in this paper. The Spearman correlation test results of the input-output variables are shown in Table 1.

From the results in Table 1, all of the correlation coefficients between variables are positive. Except for the correlation coefficient between labor and SO_2, all of the correlation coefficients pass the Spearman correlation test at the 0.01 level. The input-output variables conform to the principle of homotropy in the test, which is reasonable and can be further analyzed.

5. Calculation and Analysis

5.1. ETFEE and Its Decomposition. The ETFEE values of the BTH region from 2001 to 2016 are figured out based on the nonradial and nonoriented Window-SBM-undesirable model. Since the window technology can result in multiple values of ETFEE in a year, the average of these ETFEE values

Table 1: Spearman correlation test results.

Correlation coefficient	GDP	SO_2
Capital	0.8896**	0.4427**
Labor	0.9263**	0.0101
Energy	0.7870**	0.6035**

**A significant correlation at the 0.01 level (two-tailed).

is taken as the ETFEE value of the evaluated DMU in the corresponding year. The calculation results are shown in Figure 1.

As can be seen from Figure 1, ETFEE in Beijing, Tianjin, and Hebei shows a fluctuating growth trend from 2001 to 2016. Over the same period, Beijing and Tianjin consistently outperformed Hebei. In addition, the value of the ETFEE in Beijing was generally higher than that in Tianjin (except for 2003, 2009, 2010, and 2012). From 2009 to 2010, ETFEE values in the three provincial regions all showed a temporary small decline, which may be related to the economic crisis in 2008. In 2008, the USA subprime mortgage crisis triggered a financial tsunami, and it afterwards developed into a global economic crisis. Due to the lag effect, a local minimum of the energy efficiency curve occurred in 2010. Subsequently, ETFEE rebounded under the stimulus of a series of relevant policies issued by the central government. The ETFEE curves of Beijing and Tianjin both had small fluctuations during the study period, but in the long run, they were still on the rise, and the ETFEE values of the two cities were basically close. The ETFEE in Hebei had been growing steadily from 2001 to 2016, but it was lower than that in Beijing and Tianjin. The superior geographical location makes Hebei benefit from the development of Beijing and Tianjin, which helps to gradually raise the ETFEE of Hebei. The overall average ETFEE value in the BTH region was higher than that in the Hebei province, but lower than that in Beijing and Tianjin. Obviously, ETFEE in the BTH region is pulled down due to the low values in Hebei. It can be seen that the ETFEE in the Hebei province still has great room for improvement.

The traditional development mode, with characteristics of high energy consumption, high emissions, and high pollution, was facing severe challenges. All these problems can be attributed to the energy consumption structure. Some traditional manufacturing industries, such as iron and steel industry and cement industry, had serious overcapacities, wasting a lot of energies but reaping few benefits. After 2012, the Chinese government carried out supply-side structural reform. During this reform, BTH region shut down a large number of energy-intensive enterprises such as cement, glass, and steel enterprises. That is why the ETFEE curve of Hebei shows a rapid growth trend in this period. Later, in order to cope with the severe haze condition, Beijing, Tianjin, and Hebei implemented policies such as "coal-to-gas" and "even and odd plates." As a result, the primary energy consumption was greatly reduced, especially in Beijing, the capital of China, where coal was banned in an all-round way. The implementation of a series of measures has indeed raised the ETFEE in the BTH region, which shows that adjusting the energy structure and reducing the primary energy consumption have a certain positive effect on improving energy efficiency.

The ETFEE values in the BTH region are decomposed by the FGNZ decomposition method, and the results are shown in Table 2.

From Table 2, we know that the average technical efficiency (TE) of Beijing, Tianjin, and Hebei Prov. is 1.0170, 0.9708, and 0.8935, respectively. That is to say Beijing is DEA-efficient, while Tianjin and Hebei Prov. are non-DEA-efficient. From the perspective of pure technical efficiency, the average pure technical efficiencies of Beijing, Tianjin, and Hebei Prov. are 1.1192, 1.1496, and 0.9553, respectively. The technological progress rates of Beijing and Tianjin are 11.92% and 14.96% relatively, while that of Hebei Prov. is −4.47%. Most of the annual pure technical efficiencies of Beijing and Tianjin in the past years are more than 100% (except for 2004, Beijing), while those of Hebei Prov. are almost less than 100% (except for 2010, 2014–2016). The conclusion shows that the pure technical efficiency of Hebei Prov. is obviously lower than that of Beijing and Tianjin, which is similar to the conclusion of Liang et al [35]. In terms of scale efficiency, Beijing, Tianjin, and Hebei Prov. are 97.75%, 86.60%, and 93.81% respectively, which indicates that the energy industry scale in these three provincial areas could be further optimized.

5.2. ETFEE and Energy Structure. China relies mainly on fossil energy, and its energy consumption structure is dominated by coal, while petroleum and natural gas account for relatively small proportions. Therefore, in this paper, the consumption ratios of coal, petroleum, and natural gas are selected to represent the energy consumption structure and are recorded as SCOAL, SOIL, and SGAS, respectively. In order to eliminate heteroskedasticity between variables, the logarithms of variables are calculated and expressed as LNETFEE, LNSCOAL, LNSOIL, and LNSGAS. The ADF unit root test and cointegration test show that LNETFEE, LNSCOAL, LNSOIL, and LNSGAS are all first-order integral series. There are long-term cointegration relationships between the ETFEE and the proportion of coal, petroleum, and natural gas. In this paper, the impulse response function is used to describe the 15-period detailed state, and the long-term and short-term relationships between LNETFEE, LNSCOAL, LNSOIL, and LNSGAS are further investigated. In Figures 2–4, the horizontal axis represents the lag periods of impact action, which is 15, and the vertical axis represents the response caused by the fluctuation of the energy structure. The solid line is the impulse response function, and the dotted lines are the deviation intervals plus or minus twice standard deviations.

Figure 2 shows the dynamic response process of LNSCOAL under an impact of LNETFEE. Facing an impact of ETFEE, SCOAL was 0 in the 1st period, then began to rise slightly, and gradually declined after reaching the maximum in the 3rd period. It indicates that SCOAL has a positive effect on the ETFEE in the short term. However, in the long run, SCOAL will have a negative effect on the ETFEE, and this negative effect will expand over time. Therefore, China's

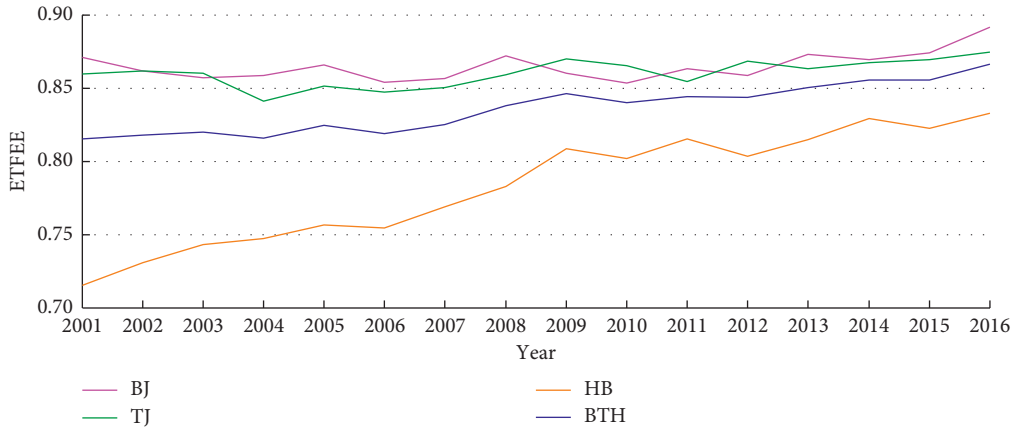


Figure 1: ETFEE in the BTH region.

Table 2: Decomposition results of the ETFEE.

Year	Beijing			Tianjin			Hebei		
	TE	PTE	SE	TE	PTE	SE	TE	PTE	SE
2001	1.0941	1.1192	0.9775	0.9735	2.0133	0.4835	0.7662	0.7793	0.9831
2002	0.9206	1.0110	0.9106	1.0007	1.3392	0.7473	0.7786	0.8445	0.9220
2003	0.9051	1.0601	0.8537	0.9189	1.1465	0.8015	0.7945	0.8758	0.9072
2004	0.8685	0.9783	0.8878	0.8800	1.0677	0.8243	0.8139	0.8647	0.9412
2005	0.9156	1.0341	0.8854	0.8907	1.0438	0.8533	0.8602	0.9072	0.9482
2006	0.9422	1.0654	0.8843	0.9032	1.0575	0.8541	0.8756	0.9208	0.9509
2007	0.9976	1.0820	0.9220	0.9278	1.0425	0.8899	0.9067	0.9611	0.9434
2008	1.0073	1.1778	0.8552	0.9830	1.0756	0.9139	0.9109	0.9786	0.9308
2009	0.9458	1.0272	0.9207	1.0686	1.1676	0.9152	0.9030	0.9262	0.9750
2010	1.0345	1.0890	0.9499	1.0599	1.1475	0.9237	0.9761	1.0267	0.9508
2011	1.0108	1.1210	0.9017	0.9471	1.0369	0.9134	0.8740	0.8990	0.9722
2012	0.9939	1.0218	0.9727	0.9371	1.0254	0.9139	0.8581	0.9078	0.9452
2013	1.0649	1.0927	0.9746	0.9542	1.0135	0.9415	0.9336	0.9891	0.9439
2014	1.0493	1.0652	0.9850	1.0013	1.0371	0.9655	0.9895	1.0598	0.9336
2015	1.0847	1.1141	0.9737	1.0380	1.0868	0.9551	1.0107	1.0785	0.9372
2016	1.5151	1.5379	0.9852	1.0486	1.0923	0.9600	1.0445	1.2658	0.8252
Mean	1.0170	1.1192	0.9775	0.9708	1.1496	0.8660	0.8935	0.9553	0.9381

TE, PTE, and SE indicate technical efficiency, pure technical efficiency, and scale efficiency, respectively.

coal-based energy consumption structure needs to be adjusted urgently to reduce the negative effect, thus raising ETFEE.

Figure 3 shows the dynamic response process of LNSOIL under an impact of LNETFEE. Facing an impact of ETFEE, SOIL continued to rise at the beginning and reached the peak in the 4th period, but then began to decline and reached the bottom in the 15th period. As time went by, the role of SOIL in improving the ETFEE gradually weakened and even had a negative effect on the ETFEE. At present, the proportion of China's oil consumption is still very low, so increasing the proportion of oil consumption in the short term will help to improve the ETFEE. In the long run, the moderate proportion of petroleum consumption should be maintained, so as to avoid negative effect on the ETFEE.

Figure 4 shows the dynamic response process of LNSGAS under an impact of LNETFEE. Facing an impact of ETFEE, SGAS showed steady upward at the beginning and increased significantly after the 8th period, which had a positive effect on the ETFEE as a whole. It can be seen that vigorously developing natural gas and increasing the proportion of natural gas in energy consumption will play a significant role in enhancing the ETFEE.

5.3. Energy Inefficiency. The calculation results show that there exist energy inefficiencies in the BTH region, as shown in Figure 5.

Figure 5 shows that energy inefficiencies in Beijing, Tianjin, and Hebei Prov. have been improving. From 2001 to 2016, the energy inefficiency curves are a trumpet-like shape. In 2001, the energy inefficiency of Hebei Prov. was close to 50%, and then the energy inefficiency curve dropped sharply year by year. In other words, although Hebei's energy efficiency had been improved gradually over the study period, there is still much room for improvement. Beijing's energy efficiency dropped from just over 10 percent in 2001 to about 5 percent in 2016, and Tianjin's dropped from nearly 15 percent to about 5 percent. Tianjin

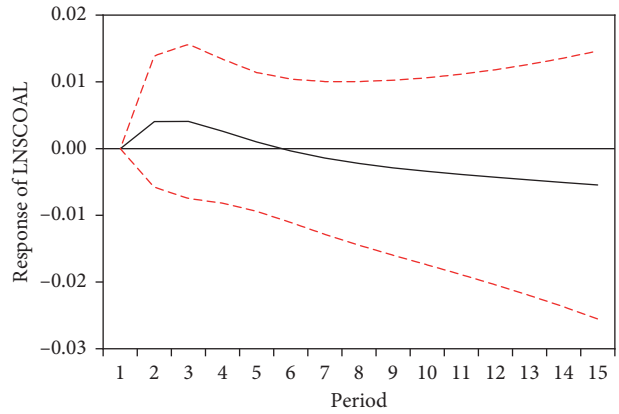


Figure 2: Impact response curve of LNSCOAL.

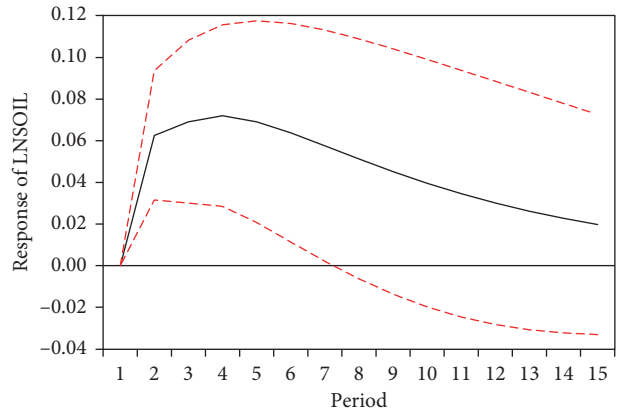


Figure 3: Impact response curve of LNSOIL.

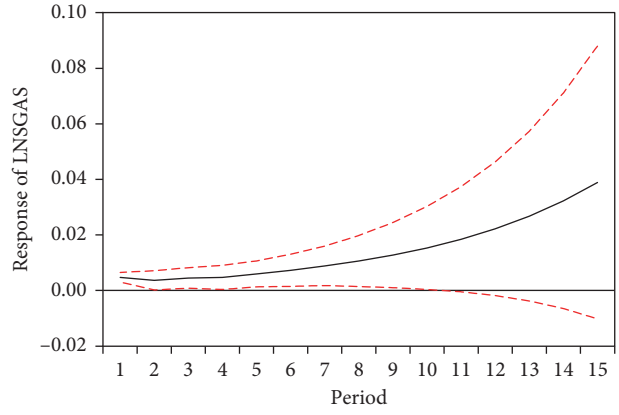


Figure 4: Impact response curve of LNSGAS.

and Beijing had similar energy efficiency levels since from 2005 to 2016 (except for 2012). It can be seen that Beijing and Tianjin are more energy-efficient than Hebei. Although the energy inefficiency curve of the BTH region shows a downward trend year by year, the overall energy inefficiency is still high, mainly due to the low energy efficiency of the Hebei province. Therefore, under the guidance of the integrated development strategy of the Beijing-Tianjin-Hebei region, it is necessary to further improve Hebei's energy efficiency so as to reach the overall energy efficiency of the BTH region.

5.4. SO_2 Reduction Potential. From the calculation results, the emission reduction potential of sulfur dioxide in the BTH region can be seen, as shown in Figure 6.

In 2001, the SO_2 reduction potential in Hebei Prov. was 82.46%, and then the potential curve fluctuated downward

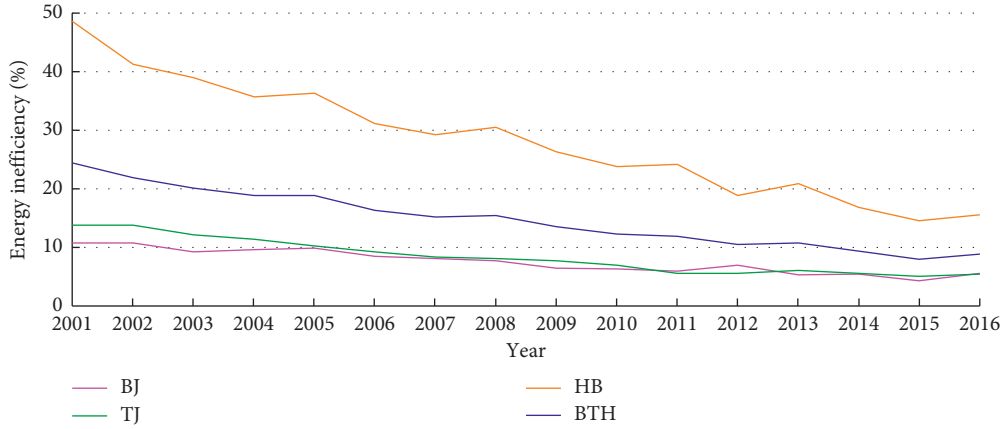


Figure 5: Energy inefficiency.

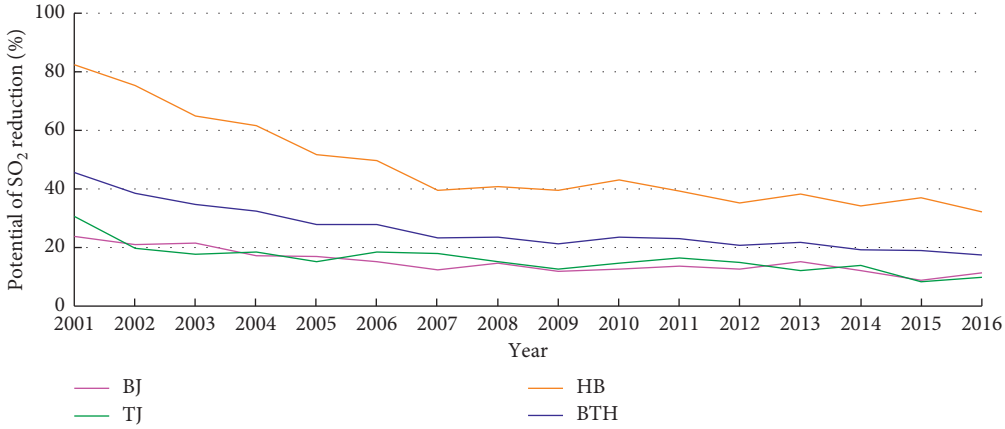


Figure 6: The potential of SO_2 reduction.

year by year. In 2010, the SO_2 emission reduction potential was 43.05%, and by 2016, it was close to 5%. The SO_2 emission reduction potentials of Beijing and Tianjin fluctuate between 10% and 20%, indicating that the SO_2 emissions of Beijing and Tianjin have been controlled to a certain extent. While Hebei's SO_2 reduction potential fell by about 50 percent during the study period, there is still plenty of room to reduce emission compared with Beijing and Tianjin. The SO_2 reduction potential in the BTH region dropped from 45.64% in 2001 to 17.48% in 2016, an overall reduction of nearly 28%. The SO_2 emission reduction in the BTH region is mainly due to the promulgation and implementation of strong energy conservation and environmental protection policies, such as the strict ban on coal in an all-round way, the improvement of oil quality, the large-scale use of clean energies, and the evacuation of polluting enterprises in Beijing and Tianjin.

6. Conclusions

Based on the unique characteristics of the BTH region and considering the undesirable output, this study establishes a nonradial and nonoriented Window-SBM-undesirable model to measure the ecological total-factor energy efficiency (ETFEE). Window data processing technology makes the data used in this paper conform to the basic empirical rule of DEA application. The Spearman correlation test shows that the input and output variables selected in this study have strong correlation and can meet the analysis requirements. From the calculation results and analysis, we can see that the developments of the ETFEE in Beijing, Tianjin, and Hebei are unbalanced. The ETFEE of Hebei Prov. has been significantly lower than that of Beijing and Tianjin. The average pure technical efficiencies of Beijing and Tianjin reach 1.1192 and 1.1496, respectively, both higher than Hebei's 0.9553. The pure technical efficiency of the Hebei province needs to be improved. From the perspective of scale efficiency, the energy structures of Beijing, Tianjin, and Hebei Prov. can be further improved. In addition, the impulse response results between the energy structure and the ETFEE show that the proportions of coal and oil will have negative effects on the ETFEE, while that of natural gas will have a positive effect over time.

The Chinese central government has put forward the integrated development strategy of the BTH region and upgraded it to a national strategy. Unremitting efforts should still be paid to increase the ecological total-factor energy efficiency (ETFEE) in the BTH region, and some actions are taken from the following aspects:

(1) Strengthen the technology transfer and improve the pure technical efficiency of Hebei Prov. It is found that the technical efficiency of Hebei Prov. is significantly lower than that of Beijing and Tianjin. As two international metropolises, Beijing and Tianjin obviously have unique technological superiorities. Under the integrated development strategy of the BTH region, policies should be made to give Beijing and Tianjin incentives to transfer technologies to Hebei Prov. In fact, for the overall harmonious development of the BTH region, Beijing and Tianjin should also try their best to improve the pure technical efficiency of Hebei Prov. as soon as possible.

(2) Work together to optimize the energy supply structure. On the one hand, Beijing and Tianjin should further insist on the utilization of clean energies, such as solar energy, wind energy, and natural gas. On the other hand, these two metropolises should help Hebei Prov. to optimize its energy supply structure. If possible, Beijing and Tianjin could subsidize Hebei Prov. for more clean energies as alternatives to coal. These three areas could coordinate the coal-to-gas supply in order to reduce the total coal consumptions in the BTH region, especially the coal consumption in Hebei Prov.

(3) Jointly develop and apply new clean and low-carbon energy technologies. At the national level, these three areas should take full advantage of the national energy supply-side reform, actively enlist national, financial, and policy support for the development of clean and low-carbon energy technologies, and gradually optimize the energy supply structure. At the local government level, Beijing, Tianjin, and Hebei Prov. should strengthen cooperation in the field of jointly developing and applying new clean and low-carbon energy technologies. For example, jointly develop the new clean coal combustion technology. As we all know, the BTH region is rich in coal, less oil, and less gas, and coal will continue to account for the majority of energy consumption in the long run. The new clean coal combustion technology is an important way to improve energy efficiency and reduce pollutant emission. In addition, recent technological innovations in photovoltaic power generation, wind power generation, and geothermal energy utilization have great potentials in clean energy technology research and development in the future. The three areas can formulate long-term energy development plans, increase official funding for clean energy development and utilization, and guide energy-related research institutions and enterprises to shift energy technology research and production to clean energies.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

The authors are grateful for the support provided by the Hebei Social Science Fund (HB17GL068).

References

[1] B. P. Global, *Global CO_2 Emissions from Energy in 2017 Grew by 1.6%, Rebounding from the Stagnant Volumes during 2014–2016, and Faster than the 10-Year Average of 1.3%*, https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy/co2-emissions.html, 2018.

[2] F. Ren and L. Gu, "Study on transition of primary energy structure and carbon emission reduction targets in China based on Markov chain model and GM (1, 1)," *Mathematical Problems in Engineering*, vol. 2016, Article ID 4912935, 8 pages, 2016.

[3] F. Ren and L. Xia, "Analysis of China's primary energy structure and emissions reduction targets by 2030 based on multiobjective programming," *Mathematical Problems in Engineering*, vol. 2017, Article ID 1532539, 8 pages, 2017.

[4] L. Stankovic, V. Stankovic, J. Liao, and C. Wilson, "Measuring the energy intensity of domestic activities from smart meter data," *Applied Energy*, vol. 183, pp. 1565–1580, 2016.

[5] H. Yan, "Provincial energy intensity in China: the role of urbanization," *Energy Policy*, vol. 86, pp. 635–650, 2015.

[6] C. Wang, H. Liao, S.-Y. Pan, L.-T. Zhao, and Y.-M. Wei, "The fluctuations of China's energy intensity: biased technical change," *Applied Energy*, vol. 135, pp. 407–414, 2014.

[7] Y. Nagata, "The US/Japan comparison of energy intensity: estimating the real gap," *Energy Policy*, vol. 25, no. 7-9, pp. 683–691, 1997.

[8] B. Su, B. W. Ang, and Y. Li, "Input-output and structural decomposition analysis of Singapore's carbon emissions," *Energy Policy*, vol. 105, pp. 484–492, 2017.

[9] R. Román-Collado and M. J. Colinet, "Is energy efficiency a driver or an inhibitor of energy consumption changes in Spain? two decomposition approaches," *Energy Policy*, vol. 115, pp. 409–417, 2018.

[10] N. Zhao, L. Xu, A. Malik, X. Song, and Y. Wang, "Interprovincial trade driving energy consumption in China," *Resources, Conservation and Recycling*, vol. 134, pp. 329–335, 2018.

[11] H. He, C. J. Reynolds, L. Li, and J. Boland, "Assessing net energy consumption of Australian economy from 2004-05 to 2014-15: environmentally-extended input-output analysis, structural decomposition analysis, and linkage analysis," *Applied Energy*, vol. 240, pp. 766–777, 2019.

[12] D. Zha, G. Yang, and Q. Wang, "Investigating the driving factors of regional CO_2 emissions in China using the IDA-PDA-MMI method," *Energy Economics*, vol. 84, Article ID 104521, 2019.

[13] B. Lin and M. Xu, "Quantitative assessment of factors affecting energy intensity from sector, region and time perspectives using decomposition method: a case of China's metallurgical industry," *Energy*, vol. 189, Article ID 116280, 2019.

[14] X. Xie and B. Lin, "Understanding the energy intensity change in China's food industry: a comprehensive decomposition method," *Energy Policy*, vol. 129, pp. 53–68, 2019.

[15] H. Rüstemoğlu and A. R. Andrés, "Determinants of CO_2 emissions in Brazil and Russia between 1992 and 2011: a decomposition analysis," *Environmental Science & Policy*, vol. 58, pp. 95–106, 2016.

[16] Y. Shang, S. Lu, X. Li et al., "Drivers of industrial water use during 2003–2012 in Tianjin, China: a structural decomposition analysis," *Journal of Cleaner Production*, vol. 140, pp. 1136–1147, 2017.

[17] P. Fernández González, M. Landajo, and M. J. Presno, "The Divisia real energy intensity indices: evolution and attribution of percent changes in 20 European countries from 1995 to 2010," *Energy*, vol. 58, pp. 340–349, 2013.

[18] P. F. González, "Exploring energy efficiency in several European countries: an attribution analysis of the Divisia structural change index," *Applied Energy*, vol. 137, pp. 364–374, 2015.

[19] K.-H. Choi and W. Oh, "Extended divisia index decomposition of changes in energy intensity: a case of Korean manufacturing industry," *Energy Policy*, vol. 65, pp. 275–283, 2014.

[20] H. Liao and Y.-M. Wei, "China's energy consumption: a perspective from Divisia aggregation approach," *Energy*, vol. 35, no. 1, pp. 28–34, 2010.

[21] J.-L. Hu and S.-C. Wang, "Total-factor energy efficiency of regions in China," *Energy Policy*, vol. 34, no. 17, pp. 3206–3217, 2006.

[22] Y. x. He, L. f. Yang, W. h. Yang, Y. j. Wang, and Y. Zhang, "Evaluation of power input efficiency for China's typical cities based on the DEA and SFA models," *International Journal of Global Energy Issues*, vol. 32, no. 4, pp. 350–360, 2009.

[23] A. Fetanat and G. Shafipour, "A hybrid method of LMDI, symmetrical components, and SFA to estimate the distribution of energy-saving potential with consideration of unbalanced components in decomposition analysis," *Energy Efficiency*, vol. 10, no. 4, pp. 1041–1059, 2017.

[24] M.-C. Chang, "A comment on the calculation of the total-factor energy efficiency (TFEE) index," *Energy Policy*, vol. 53, pp. 500–504, 2013.

[25] J.-P. Liu, Q.-R. Yang, and L. He, "Total-factor energy efficiency (TFEE) evaluation on thermal power industry with DEA, malmquist and multiple regression techniques," *Energies*, vol. 10, no. 7, p. 1039, 2017.

[26] B. Wang, J. Zhang, and H. Zhang, "Total-factor energy efficiency and influencing factors across provinces in China in the presence of environmental regulation," *Economic Review*, vol. 4, pp. 31–43, 2011.

[27] K. Tone, "Dealing with undesirable outputs in DEA: a slacks-based measure (SBM) approach," in *Proceedings of the North American Productivity Workshop 2004*, Toronto, Canada, June 2004.

[28] R. W. Pasurka, "Multilateral productivity comparisons with undesirable outputs," *The Economic Journal*, vol. 93, no. 372, pp. 883–891, 1983.

[29] X. Wang, M. Meng, J. Liu et al., "Study on total factor energy efficiency of Beijing-Tianjin-Hebei metropolitan region under the constraint of carbon emission," *Journal of Industrial Technological Economics*, vol. 32, no. 1, pp. 11–19, 2013.

[30] B. Feng and X. Wang, "Research on energy efficiency considering the haze effect in Beijing-Tianjin-Hebei metropolitan region," *Journal of Arid Land Resources and Environment*, vol. 29, no. 10, pp. 1–7, 2015.

[31] S. Wang, "Research on the difference and the coordination of Jing-Jin-Ji's energy efficiency in perspective of green development," *Forum on Science and Technology in China*, vol. 10, pp. 96–101, 2016.

[32] W. W. Cooper, *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*, Springer Science & Business Media, New York, NY, USA, 2nd edition, 2007.

[33] J. Zhang, G. Wu, and J. Zhang, "The estimation of China's provincial capital stock: 1952–2000," *Economic Research Journal*, vol. 10, pp. 35–44, 2004.

[34] R. Färe, S. Grosskopf, and C. A. Pasurka Jr., "Potential gains from trading bad outputs: the case of U.S. electric power plants," *Resource and Energy Economics*, vol. 36, no. 1, pp. 99–112, 2014.

[35] Y. Liang, D. X. Niu, W. W. Zhou et al., "Decomposition analysis of carbon emissions from energy consumption in Beijing-Tianjin-Hebei, China: a weighted-combination model based on logarithmic mean divisia index and Shapley value," *Sustainability*, vol. 10, no. 7, p. 2535, 2018.

Hindawi
Discrete Dynamics in Nature and Society
Volume 2020, Article ID 7823971, 15 pages
https://doi.org/10.1155/2020/7823971

Research Article

An AHP-DEA Approach of the Bike-Sharing Spots Selection Problem in the Free-Floating Bike-Sharing System

Minjiao Cheng and Wenchao Wei

School of Economics and Management, Beijing Jiaotong University, Beijing, China

Correspondence should be addressed to Wenchao Wei; weiwenchao@bjtu.edu.cn

Received 25 June 2020; Revised 23 August 2020; Accepted 3 September 2020; Published 18 September 2020

Academic Editor: Chi-Hua Chen

The Free-Floating Bike-Sharing system (FFBSS) connects users to public transit networks and is an important component of the "last-mile" transport network. However, the rapid development of the FFBSS in China has significantly increased the local municipal workload and deteriorated the public transport. To mitigate these negative impacts, the Chinese government has launched a pilot project of electronic fence spots, by selecting several bike-sharing parking spots from the existing ones to set up the virtual stations. Compared to the traditional public shared bicycles with fixed stations, the flexibility of choosing parking spots could dynamically cater the fast-changing traffic environment and facilitate the renting and returning for users and, therefore, render a more sparse and complicate parking spots network. In this paper, we study the location selection of bike-sharing parking points as a multidimensional problem, which considers not only the interests of users and stakeholders but also the environment and safety issues. We propose a multicriteria decision-making (MCDM) model including the analytic hierarchy process (AHP) and the weight-restricted data envelopment analysis (DEA) method to evaluate and determine the optimal bike-sharing parking points. Since the additional weight restrictions in the DEA method might lead to infeasible solutions, we introduce the weight restrictions feasibility theorem to avoid such infeasibility in the proposed weight-restricted DEA model, which has not been thoroughly studied in the literature. Specifically, a hyperplane adjusting model is developed to adjust the infeasible results. In the computational study, we evaluate 36 current parking spots in three regions in Beijing, China, to verify the rationality of the combined approach and put forward some managerial suggestions for this pilot project in Beijing, China.

1. Introduction

The bike-sharing system (BSS) is a short-time rental service system to provide customers with bikes for shared uses [1]. As a cheap, convenient, and emission-free travel method, it is suitable for small distances commuting, especially for the first- and last-mile problem in multimodal transportation in urban regions [2–4]. Although the bike-sharing programs with the earliest generation in 1965 (White Bikes, Amsterdam) have been existing for more than half a century, they recently experienced a sharp increase in their worldwide popularity along with the advent of the free-floating bike-sharing system (FFBSS). As pointed by Liu et al. [5] and Sun et al. [6], the FFBSS has more flexible parking spots locations than the traditional BSS and only requires a small area for parking, such as sidewalks, commercial plazas, residential vacancies, and other small pieces of the city. Besides, the lower construction cost and small service scope render a dense distribution of bike-sharing parking spots [7]. But, the scale-up of the bike-sharing programs may generate negative trade-offs, including abuse of public spaces and undermining long-term environmental and social sustainability in cities [8]. In the recent years, the Chinese government has launched a pilot project of electronic fence points to mitigate these negative impacts. The electronic fence points are similar to the traditional parking stations but are literally virtual stations using the Bluetooth Road Stud technology to restrict the parking regions and belong to the public infrastructures, which allow different brands of shared bikes to park in.

The selection of suitable parking spots is one of the key factors for the success of this pilot project by reducing the number of parking spots to a proper manageable level while

satisfying the users' requirements. However, attention should be paid to the rationality of the site selection to minimize the impact of the reduction of parking spots on users. Therefore, the method prioritizing suitable locations for bike-sharing parking spots is desirable. The selection of bike-sharing parking spots is one of the facility location problems, which refers to the resource allocation optimization to choose the best location of parking spots for different objectives and constraints. Many scholars have made different attempts to solve this issue. For instance, the traditional location approach such as the P-median model and maximum coverage model [9–11], the mixed-integer linear program (MILP) [6, 12, 13], the robust optimization method [14, 15], Data Mining [16], and GIS technique [17]. Although these methods have tackled the facility location problem efficiently to some extent, they are not suitable in a multicriteria objective optimization environment, i.e., the electronic fence selection problem, which involves operational costs, convenience, safety, and municipal management issues in the objective function. Furthermore, the selection of the appropriate electronic fence points for the FFBBS is a complex problem and requires an extensive evaluation process because it is not trivial to develop a selection criterion that can precisely describe the preference of one location over another. Besides, we will only consider deploying the electronic fence on the existing parking spots, rather than on a newly selected spot. Therefore, the selection of these sites could be viewed as a multicriteria decision-making (MCDM) problem and tackled in the MCDM framework [2, 18, 19].

The contribution of this paper is fourfold. Firstly, we propose an advanced criteria system for selecting the optimal bike-sharing parking sites with a comprehensive consideration of relevant factors, including safety, the convenience of users, the interest of companies, and the managerial difficulties, for municipal. Secondly, we justify that a hybrid approach of weight-restricted DEA and AHP with the MCDM framework could be used to solve this selection problem. Up to our knowledge, this is also the first time such a method is used in the bike-sharing spots selection problem. Thirdly, we conduct a feasibility test on the proposed approach which is not seen in pioneer studies and develop an adjusting model to modify the method. Finally, a comparative case study is provided, where we evaluate 36 current parking spots in three regions (namely, Zhongguancun Science Park, the Guomao CBD, and the Liyuan Subway Station) in Beijing, China. The comparison between the results of the basic BCC model and weight-restricted DEA is also conducted to justify the hybrid method's superiority. It not only verified the rationality of the method but also put forward some management suggestions for the pilot work in Beijing, China.

This paper is organized as follows: Section 2 reviews the literature in the evaluation and selection of bike-sharing parking spots. Section 3 introduces the advanced criteria system, the method which combines the weight-restricted DEA evaluation model with an AHP selection process and the weight restrictions feasibility theorem and the hyperplane adjusting model. A case study of several current bike-sharing spots in Beijing, China, is conducted in Section 4. The conclusions and future research directions are summarized in Section 5.

2. Literature Review

As mentioned by Garcıa-Palomares et al. [10], the location selection of bike-sharing parking spots can be formulated as an MCDM problem, and MCDM models have been developed to assess bike-sharing stations. Kanjanakorn and Piantanakulchai [20] applied the AHP to rank suitable locations of the bike-sharing station by considering factors from the planners' and users' perspectives, such as types of land, amount of available space, accessibility to the main bike route, and walkability to destinations. Kabak et al. [2] combined the AHP and MOORA techniques with the geographic information system (GIS) to assign the priorities and rank bike-sharing stations by evaluating the conflicting criteria, which included the population, recreation areas, cycle line, and public transport networks. Among the existing MCDM techniques, i.e., the AHP, analytic network process (ANP), the technique for order of preference by similarity to ideal solution (TOPSIS), multiobjective optimization based on ratio analysis (MOORA), and DEA, the key to success is the proper decoupling points of method and reality; therefore, researchers tend to apply these methods either in a standalone fashion or in a combined way regarding the differences of real-world applications [21, 22]. Javadi et al. [19] attempted to find the best locations for bike-share stations in the city of Isfahan using both mathematical programming and MCDM techniques. They selected four main factors of bike-sharing location such as closeness to the bicycle path, transportation and networks, and demand and used type and applied AHP to weigh these appropriate criteria. This criteria system included the measurement of the convenience and the operational cost. Then, the final weights of the proposed locations have been calculated through a simple additive weighting method. For the determination of the multicriteria system, besides cost and convenience considerations, attention to other social factors, such as safety factors and the managerial level in a station, is also crucial [23]. However, a comprehensive consideration of all these factors, to the best of our knowledge, is still rare in the literature. This paper intends to fill this gap.

Among the extensive methods to evaluate the relative efficiencies of a set of decision-making units (DMUs), DEA is an objective approach using no parametric linear programming-based techniques and has been widely used in location selection problems, i.e., the multiple-objective layout problem completely depending on the objective data [24]. However, a subjective orientation is also important in a siting analysis whenever uncertainty, human values, or other subjectively oriented elements may play a vital part in the decision [25]. Therefore, it is desired to integrate the decision-makers' (DM) subjective judgment into the bike-sharing parking spots selection problem. The AHP approach, developed by Saaty [26], is effective when the decision-making process involves structuring multiple-choice criteria into a hierarchy to assess the "importance" weights to all the relevant decision elements,

either objective or subjective. Yang and Kuo [27] integrated the DMs' preference into the evaluation framework by using the AHP technique to measure the qualitative factors; therefore, it leads to more efficient identification of selection criteria, weighting, and analysis.

The combination of the AHP and DEA methods integrates the subjectivity judgments and the objective selection and, therefore, renders a more accurate and reasonable result. The method of using the AHP technique to incorporate preference information into the DEA model is widely used in many situations, such as the selection of a manufacturing system for a manufacturing organization [28], the evaluation of materials [22], and the selection of optimal suppliers for the oil industry [29]. In the research area of facility location, Korpela et al. [30] combined the AHP and DEA to select the location of warehouses. Khadivi and Fatemi Ghomi [31] adopted the ANP and DEA approach in two stages to evaluate the solid waste facilities' locations. Mohajeri and Amin [18] introduced the railway station site-selection problem as a hierarchy model consisting of four levels and, then, incorporated DEA to aggregate the AHP global priorities. In most facility location researches, the AHP model is only used to generate qualitative input figures for the DEA model, where the different importance of the criteria to the system is not properly considered. To overcome this issue, we argue that the weight-restricted DEA method which has been successfully applied in many scenarios is also promising in handling the bike-sharing spot selection problem. The corresponding justification is presented in Section 3. The only paper we found that focused on a similar topic (facility location) is [32]. In their work, the authors proposed a consensus-making method, which is based on a combination of the AHP and the assurance region (AR) model of DEA, to reach a group decision for the relocation problem of government agencies out of Tokyo. In their paper, the appropriate weights of criteria were calculated through an AHP procedure, and the additional weight constraints for the DEA model were generated by the AR technique. On the one hand, the weight restrictions of multipliers can avoid the extreme values which violate the preferences of DMs. On the other hand, this hybrid method can increase the discrimination between DMUs when the number of inputs and outputs is much larger than the number of DMUs [33]. It should be noted that additional constraints might lead to infeasible solutions to the linear programs in the DEA model [34–36]. This paper extends Takamura and Tone's [32] work by introducing an adjusted model to revise the infeasible solutions and generate reasonable results. The detailed procedure is available in Section 3.

Above all, the existing research of the bike-share parking spots evaluation criteria mainly focused on the FFBBS's productivity and users' demand. In this paper, we propose a more comprehensive criteria system to extensively describe the complex and dynamic nature of the electronic fence points selection problem. Four major criteria and eight subcriteria are identified in this system. Due to the complexity in considering the qualitative information and discriminatively quantifying the criteria, we apply a hybrid method of the AHP technique and weight-restricted DEA within the MCDM framework to streamline the evaluation process of bike-sharing parking spots locations. The AHP approach is used to assess the qualitative indicators, measure the DM's preferences among these indicators, and finally, generate additional weight restrictions to be added in the DEA method. A hyperplane adjusting model is also developed to adjust the infeasible results in the hybrid approach.

3. Methodology

3.1. Criteria System of MCDM. The proper identification of the criteria in the evaluation system should reflect the real-world operational scenarios and is vital to the results' rationality [4]. The evaluation criteria system should accurately assess the efficiencies of the existing bike-sharing parking points and figure out the most promising ones. An advisory board is organized to determine the criteria, which consists of three transportation engineers, two municipal planners involved in the pilot project of electronic fence points, two experts from the China Bicycle Association, three operators, and five users (the details are explained in Table 1).

After the interview with the advisory board, the factors affecting the location of bike-sharing parking spots are grouped into four main criteria, specifically, accessibility to the destination (C1), walkability to the transportation hub (C2), the flexibility of bicycle renting and returning (C3), and the interest of the bike-sharing company (C4). We also reviewed the relevant literature to further examine the most influential factors of selecting bike-sharing parking spots (see Table 2). The overall explanation of the criteria is in Table 2.

(i) Accessibility to the destination (C1) includes the proximity to the destination entrance (C11) and the condition of the pedestrian walkway between parking points and destinations (C12). Many papers have considered the factor of C11; however, C12 is ignored by most researchers (see Table 2). The only research considered C12 [20] tackled a similar problem and argued that the pedestrian condition is an important factor that should not be neglected in cities with high population density. Good condition of the pedestrian walkway means that the passengers would get to their destinations in a fast and safe manner, while a deteriorated pedestrian walkway causes inconveniences to the passengers and prevents them from using the parking spots. Therefore, the condition of the pedestrian walkway is considered in the criteria system.

(ii) Walkability to the transportation hub (C2) consists of subcriteria such as proximity to the nearest bus stations (C21), proximity to the nearest parking spots (C22), and proximity to the nearest parking spots (C23). In large cities, bike-sharing is an important component of the "last-mile" transport to connect users to the public transit networks. It is not surprising that the first two criteria are considered in most relevant papers (as shown in Table 2). The

Table 1: The members of the advisory board.

Members	Number of experts	Working experience (years)
Transportation engineers	3	10–15
Municipal planners	2	20
Experts from the bicycle association	2	5–8
Operators of the bike-sharing company	3	5–8
Bike-sharing users	5	—

Table 2: Considered criteria and relation with the literature.

	C1		C2		C3		C4	
	C11	C12	C21	C22	C23	C31	C32	C41
Garcıa-Palomares et al. [10]	√		√	√		√		√
Kanjanakorn and Piantanakulchai [20]	√	√				√	√	
Frade and Ribeiro [9]			√	√				
Murphy and Usher [37]	√		√	√				
Cetinkaya [38]	√		√	√		√		√
Kabak et al. [2]	√		√	√		√		√
This paper	√	√	√	√	√	√	√	√

distance to the nearby parking spots represents the service radius of the parking spots and also indicates the time a customer should expect to find a parking spot. Besides, the setup cost of a parking spot renders high operational expenses for intensive parking spots deployment. Therefore, C23 is also an important indicator explaining the tradeoff between service quality and capital investments and deserves its right to be considered in the criteria system.

(iii) Flexibility of bicycle renting and returning (C3) is a factor measuring the conveniency for users to retrieve the bicycles. The proximity to the bike lane (C31) denotes the distance to the bike lane and the corresponding riding quality. It is often used to describe the safety and the convenience of riding a bike. The amount of available space (C32) indicates the size of the area and the number of bicycles which can be parked. It should be noted that, during the rush hours (i.e., 7–9 am), insufficient parking space might force a customer to find an undesirable place, i.e., a far-way parking spot, pedestrian, or shrubs beside the road, to park his/her bikes. The imbalances between bike-holding amount and dynamic bike requirements in different stations often lead to a shortage of bikes or lack of parking space. Besides, a parking spot of the large size normally needs a large amount of investment and much managerial effort. All these issues require a dedicated study on determining the size of each parking spot. Therefore, C32 is included in our study as well.

(iv) The interest of the bike-sharing company (C4) is reflected by the population density (C41). For private businesses, a high turnover rate of bikes is vital to maintain a healthy business. It is reasonable to locate the parking spot in the area with a high volume of population.

The overall criteria system is summarized in Figure 1. Considering the different priorities of criteria, an AHP approach is applied to determine the weight coefficient of each criterion. The methodology of the AHP in this study follows Subramanian and Ramanathan's [39] research, and we put the detailed procedure in Appendix A.

3.2. Weight-Restricted DEA Models. For basic BCC models (shown in Appendix B), the multipliers of inputs (u) and outputs (v) are derived from the data instead of being fixed in advance. In the AHP tree for bike-sharing parking spots selecting, the importance of each input indicator is not equal. Without the restrictions of multipliers, it may cause tremendously large or small weights for some criteria whereas these extreme values are undesirable considering the DMs' preference. Besides, a large number of input and output indicators in the evaluation framework would cause an overestimation of the parking spot's efficiency. The usage of weight-restricted DEA is not only to incorporate the preference information into the basic DEA method but also to diversify the discrimination between DMUs when the number of inputs and outputs is much larger than the number of DMUs. Stawicki and Lawrence [40] compared the DEA with and without weight restrictions, respectively. Finally, they concluded that when the weighting given to the various inputs and outputs is forced into a narrow band across all factors, their efficiency drops off significantly.

There are many methods for incorporating DMs' value judgment in DEA assessments [41]. Allen et al. [33] concluded that the incorporation approaches are very application-driven: in real-world applications, there have emerged situations where some preference information is needed, and this has dictated the way it has been incorporated. We divide the methods into three categories and specify the applications using each approach, see Figure 2.

Direct weight restrictions could be the absolute upper and lower bounds on the weights proposed by Roll et al. [47].

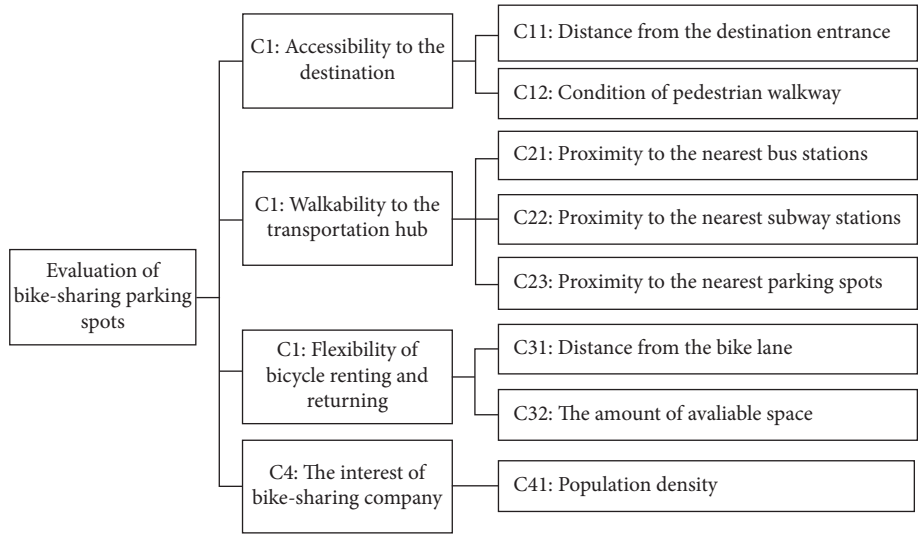


Figure 1: Proposed criteria system.

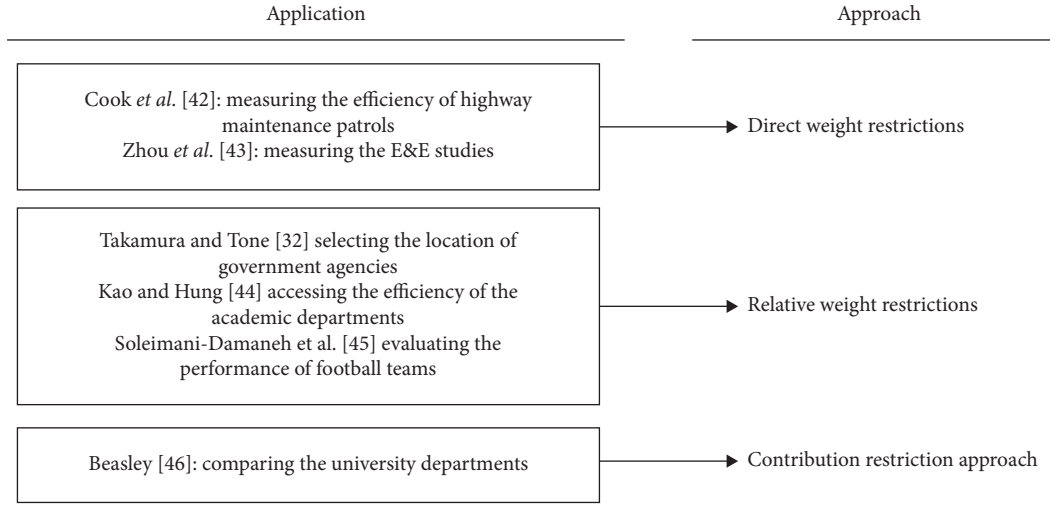


Figure 2: Approach for incorporating value judgments in DEA [42, 43].

Since the magnitude of the weights in the direct weight restrictions depends on the scale of the inputs and the outputs, it is difficult and unintuitive for a DM to find the restrictions that capture his/her preferences [48]. In the contribution restriction approach, which was firstly introduced by Wong and Beasley [49], the lower and upper bounds are placed to the contribution of each input and/or output. However, this weight-restriction scheme could lead to more additional restrictions. Due to the computational expensiveness of the contribution restriction approaches, we utilize the relative weight restrictions AR to incorporate the DMs' preferences into the location selection problem [50]. The idea of this approach is to restrict the relations of weights of the inputs and/or outputs and to exclude virtual multiplier vectors that are not reasonable from the model [32, 45, 51, 52]. In this paper, the AR is specified by the AHP method. We mainly reference the procedure of Soleimani-Damaneh et al.'s [45] paper, which calculated the ratios of other outputs/inputs weights and the weight of the first output/input. In this paper, we only consider the relative ordering or values of the multi-input factors because of the brief components of outputs, such as user satisfaction and the bike-sharing turnover rate.

We add the restrictions $Av \leq 0$ into the input-oriented BCC model (see equation (B.1) to (B.4) in Appendix B). The envelopment form (dual) of the weight-restricted DEA model is as follows:

$$\min \quad h, \tag{1}$$

$$\text{s.t.} \quad X_0 h - X\lambda + A^T \gamma \geq 0, \tag{2}$$

$$-Y_0 + Y\lambda \geq 0, \tag{3}$$

$$1^T \lambda = 1, \tag{4}$$

$$\lambda \geq 0, \tag{5}$$

where A is a $p \times n$ matrix and λ and γ are the dual vectors.

3.3. Models to Test the Feasibility. Note that adding the weight boundaries to the DEA model might lead to infeasible DEA evaluations. Chang and Wang [53] pointed out two shortcomings of AHP restrictions: the subjectivity and inconsistency of dating the AHP in DEA and the local optimality of DEA. To overcome the probabilities of infeasible DEA results, the methods to avoid inefficiency are discussed in many papers. For instance, Podinovski [48] proposed a maximum DEA model to avoid the side effects of incorporating absolute weight bounds, and it is equivalent to the classical DEA model if no additional restrictions are imposed. Lins et al. [34] proposed the weight restrictions feasibility theorem to test the feasibility of the results and, then, introduced the adjusting models to modify the results. In this paper, we utilize the weight restrictions feasibility theorem given by Lins et al. [34] to test whether the BCC models are still feasible after adding the weight restrictions.

Theorem 1. *If there exists no vector $\gamma \geq 0$ such that $A^T\gamma > 0$, then conditions $Av \leq 0$ are met and the weight-restricted DEA model is feasible.*

Theorem 2. *Instead, if there exists a vector $\gamma \geq 0$ such that $A^T\gamma > 0$, then conditions $Av \leq 0$ are not met and the weight-restricted DEA model is infeasible.*

Proof. For envelopment program equations (1) to (5), we suppose that $\forall\gamma \geq 0$ and there exists, at least, one row k of A^T which makes $(A^T)_k^* \gamma \leq 0$. For constraint (2) we have $X_{0k}h \geq X_k\lambda - (A^T)_k\gamma$. Therefore, the variable h to be minimized has the lower bound, which means the weight-restricted model is feasible. Conversely, suppose $\exists\gamma \geq 0$ such that $A^T\gamma > 0$. γ is not bounded, then $A^T\gamma$ is not bounded as well. Thus, h has no lower bound, which means this model is infeasible.

If we find one γ to make $A^T\gamma > 0$, we need to relax the weight bounds by using the adjusted weight matrix $A - A'$. The hyperplane adjusting model with the objective of finding the minimum loosening adjustments is as follows:

$$\begin{aligned} \min \quad & 1^T A' 1, \\ \text{s.t.} \quad & (A - A')v \leq 0, \\ & v, A' \geq 0, \end{aligned} \tag{6}$$

where A is the upper bound assigned to the ratio of two input indicators, A' is the adjustment matrix, and 1 is a column unit vector.

The process of the proposed methodology solving the parking spots selection problem is illustrated in Figure 3. There are five main steps in our procedure, such as the determination of criteria, the collection, construction of the database (qualitative data and qualitative data, respectively), calculating for the current parking spots using this hybrid AHP-DEA approach, testing for the feasibility of our models and adjusting the infeasible results, and ranking for the alternative parking spots. We will use this procedure in the case study.

4. Case Study

In this section, we conduct a case study on three selected regions in Beijing, China. The proposed hybrid AHP-DEA method is applied to process the collected data. We also provide a comparison between this hybrid approach and the conventional method and managerial insights to the parking spots selection problem during the electrical fence deployment campaign.

4.1. Region Selection. Beijing is the capital city of China and is one of the largest cities in China with approximately 21.5 million inhabitants. In 2019, the average daily bike-sharing ride volume in Beijing reached to 1.272 million times, and the average daily turnover rate was 1.4 times per vehicle (Beijing Municipal Commission of Transport). The Beijing Municipal Government has launched the first batch of electronic fence points in three districts: Haidian, Chaoyang, and Tongzhou. The heat map of riding demand in the urban region is shown in Figure 4 (OFO big data). Haidian district is located in the northwest part of Beijing with a greater demand for bike-sharing in the heat map. Chaoyang district has the largest area in the downtown and is located in the east part of the urban area with a moderate riding demand. To the east of Chaoyang district is the Tongzhou district with less demand for cycling. In this study, three different regions from the abovementioned districts are selected as the testbeds of our evaluation approach (the three red circles depicted in Figure 4).

Considering the small service scope, relatively dense layout, and wide range of bike-sharing parking points, the analysis of parking spots' efficiency should focus on the riding routes in the bike-sharing system, i.e., between parking points and destinations [20]. Besides the conventional method dividing the target regions by the volume of riding demand (Large, Medium, and Small), we distinguish the regions by their surroundings as well (shopping malls, residential area, and so on). The purpose is to evaluate the effectiveness of our hybrid method and extend the boundaries of our research. These selected regions are all with high population mobility, namely, Zhongguancun Science Park, the Guomao CBD, and the Liyuan Subway Station, as shown in Figures 5–7, respectively. The features of these three small regions are described in Table 3.

In each region, 12 bike-sharing parking spots are selected (in total 36 spots). The selection of these spots is mainly based on the current bike-sharing networks. The index and location of each spot is illustrated in Figures 5–7. The AHP method is applied to calculate the weights of criteria with regard to the DMs' preference. The obtained weights are, then, incorporated into the weight-restricted DEA model to rank the parking points in each zone.

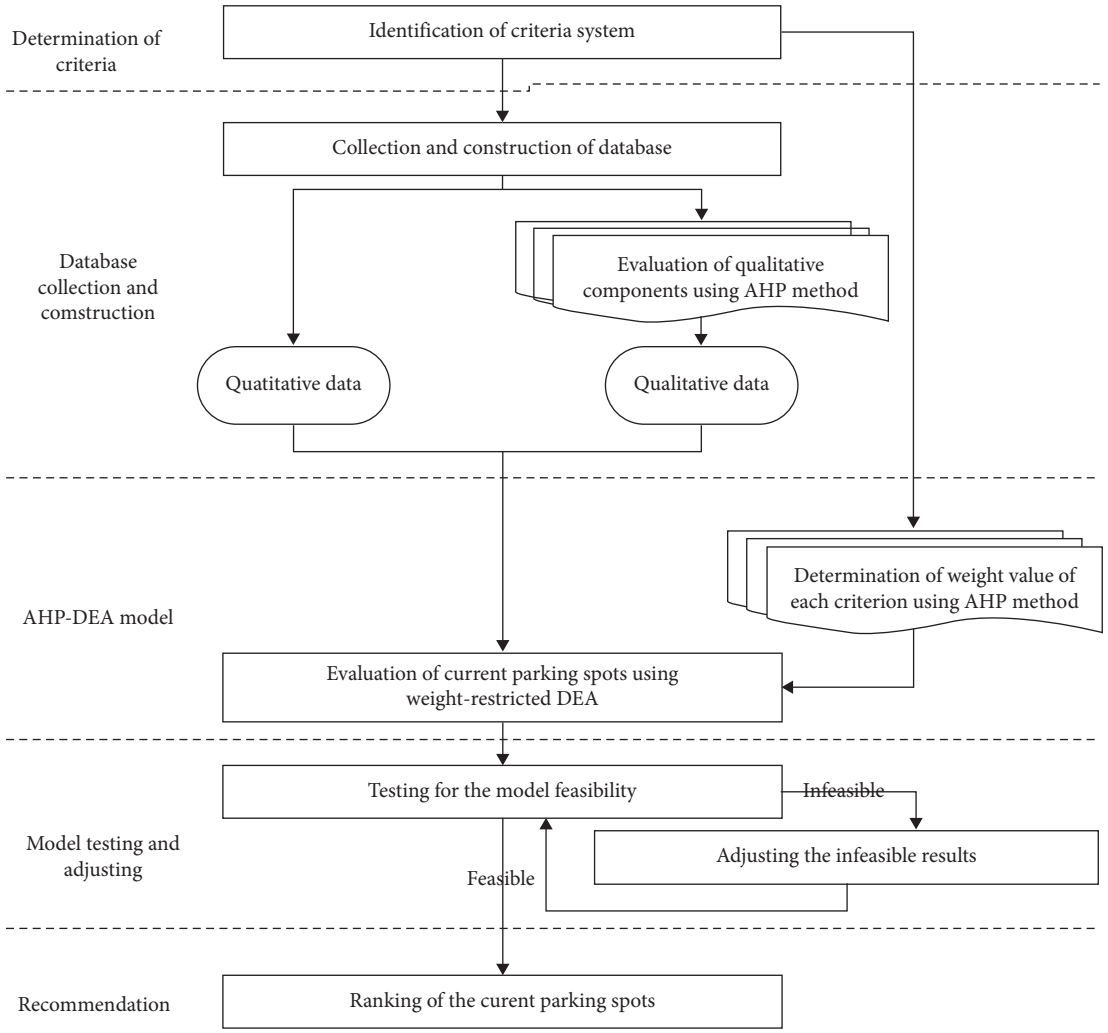


Figure 3: Methodology for the spots selection process.

4.2. Data Processing and Calculating. The data of input factors C11, C21, C22, C23, C31, and C32 are obtained from the investigation in the field. The qualitative factors C12 and C31 are assessed by the AHP technique. C41 is extracted from the Geographical Information Monitoring Cloud Platform (China). For the output factors, we consider the user satisfaction (O1) and bike-sharing turnover rate (O2). O1 is acquired based on a questionnaire survey. Note that, to supplement the interview data, additional data from a variety of sources, including business and media reports, were also utilized. O2 is the volume changes of bicycles in one specific parking point between two consecutive observation periods and is obtained from the investigation in the field.

In this study, the members of the advisory board are required to make pairwise comparisons by using questionnaire and face to face interview to determine the relative importance of the input factors. The question is like "how do criterion A and B relatively differ in importance?"[‡] Super Decisions V2.10 for Mac is used to calculate the results. Afterward, we calculate the geometric mean for the experts' opinions, and the values are collected in Table 4. The CR value is less than 0.1, which means the data comparing the criteria pairs is appropriate and does not need to be re-evaluated.

As we can observe from Table 4, the most significant factors concerned by the members of the advisory board in locating bike-sharing parking spots are the proximity to the bus stations and subway stations (both with $w = 0.2175$) and the population density ($w = 0.1706$). The proximity to the bike lane ($w = 0.0396$), the condition of pedestrian walkway ($w = 0.0460$), and the amount of available space ($w = 0.0649$) are less important factors from the aspect of FFBSS operations. According to these values, we impose the weight restrictions using the AR I weighting scheme, where weights of the other inputs are related to the weight of the first input. This procedure is also proposed in the work of Joro and Korhonen [41] and Soleimani-Damaneh et al. [45]. The resulting weight matrix A is as follows:

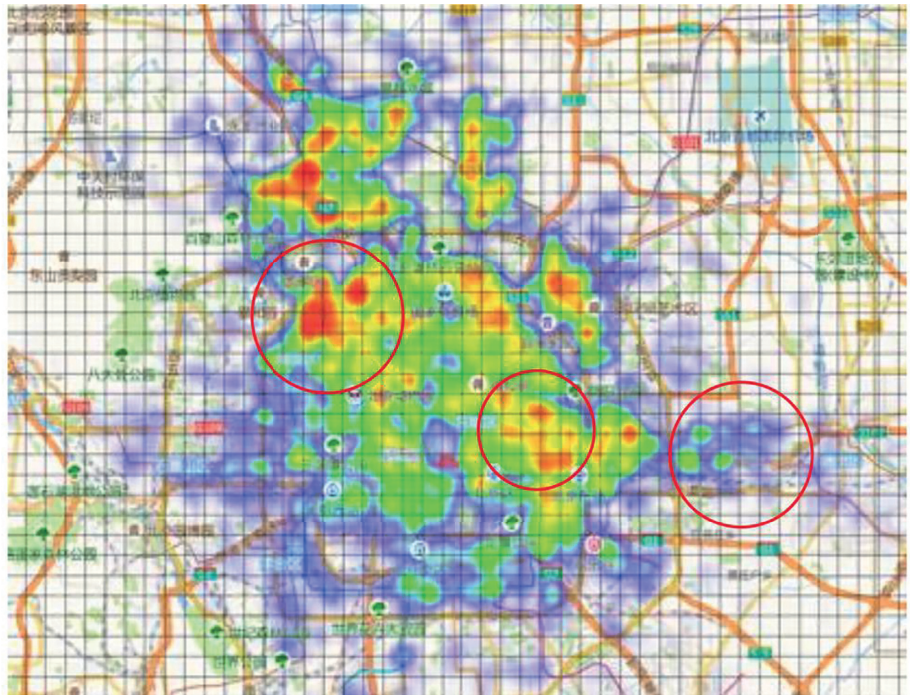


Figure 4: The heat map of riding demand in Beijing.

$$A = \begin{bmatrix} 1 & -3 & 0 & 0 & 0 & 0 & 0 & 0 \\ -2 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ -2 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 3 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}. \quad (7)$$

Adding the input weight constraints $Av \leq 0$ to the basic BCC model, we would obtain the weight-restricted DEA scores, as shown in the fourth column of Tables 5–7. The basic BCC scores are listed in the second column for the sake of comparison.

To verify the feasibility of the evaluation, we test the results using the weight restrictions feasibility theorem described in Section 3.3. We observe that there is no vector $\gamma > 0$ making $A^T\gamma > 0$, which means the result is feasible.

4.3. The Comparison of Basic BCC Results and Weight-Restricted DEA Results. In Tables 5–7, the rankings of the evaluations scores based upon the obtained efficiency scores are summarized in the third and fifth column of these three tables. It can be observed that 9 parking spots (75%) in Zhongguancun Science Park have efficiency score equal to 1, and the value for the Guomao CBD and Liyuan Subway Station is 7 (58%) and 9 (75%), respectively. This means that most points are on the efficient frontier. The overestimation of efficiency scores relies on the large number of input and output factors in the DEA framework. In the conventional evaluation procedure, the number of DMUs is normally two times more than the number of input and output factors [41]. The addition of the weight restrictions into the BCC model would decrease the efficiency scores of most DMUs, as shown in Figure 8. This is one of the advantages of weight-restricted DEA, which can increase the discrimination between DMUs when the number of inputs and outputs is much larger than the number of DMUs.

Furthermore, we can find that the efficiency average considering the DMs' preferences in the Guomao CBD (0.8867) is less than the efficiency average in Zhongguancun Science Park (0.9295) and the Liyuan Subway Station (0.9347). However, without the consideration of the criteria priority, Zhongguancun Science Park has the lowest efficiency average and the DMs' preferences have the largest impact.

The efficiency score of spot 1 changes from 1 to 0.9396. The high score in the BCC model is due to the short distance from the subway station, while it is far from the destination entrance than the other alternatives. The distance from the destination entrance is considered as one of the most important factors for the evaluation (see Table 4). The same reason explains the score changes of spot 10, which is near the gate of community and is close to the nearest parking spots. However, it is far from the subway stations and the bus stations. For spot 13–24, all of them are far away from the subway stations and bus stations.

After adding the weight restrictions to the basic BCC model, six spots have a lower efficiency score. Among them, spot 21 changes from the performance-efficient to the inefficient. That is because it has the smaller population density, which is a major influential factor for the selection of electronic fence spots. The efficiency score of spot 23 changes from the rank of 10 to the rank of 8. The small population density renders a lower score, while the short

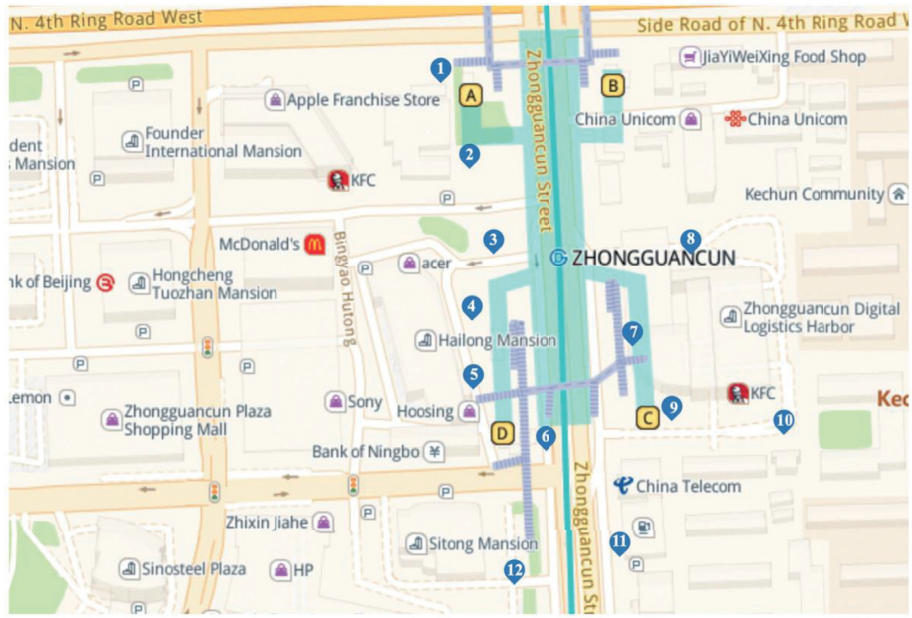


Figure 5: Twelve current bike-sharing parking spots in Zhongguancun Science Park.

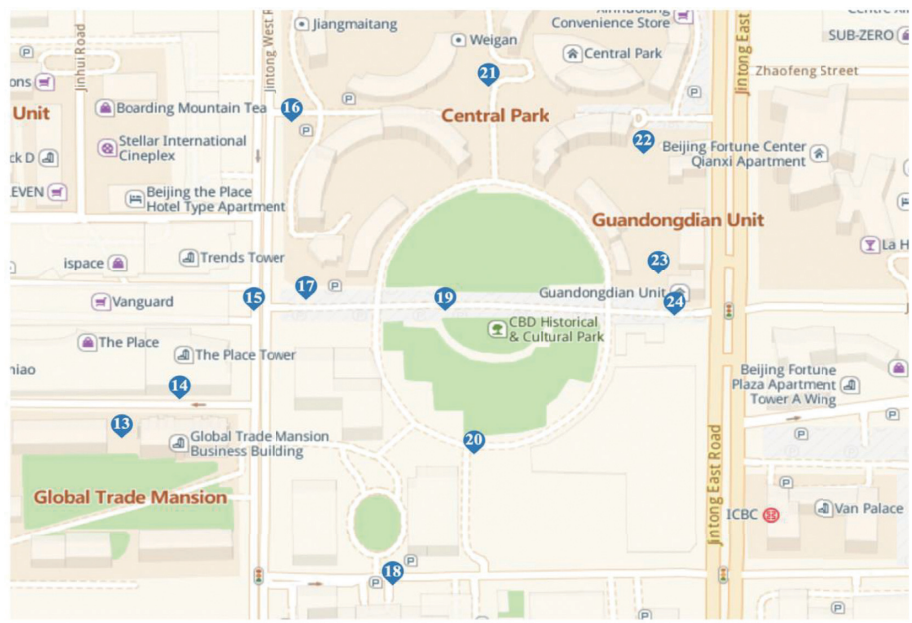


Figure 6: Twelve current bike-sharing parking spots in the Guomao CBD.

distance from the destination entrance and parking spots raises its ranking. For the efficiency scores of spot 25–36, there are only 4 spots, including spot 27, spot 30, spot 33, and spot 36, on the efficiency frontier, when different importance of the input indicators are considered. Although the spot 25 is close to the destination entrance, its rank drops significantly because of its long distance from the subway station.

Spot 4 (in Zhongguancun Science Park), spot 20 (in the Guomao CBD), and spot 26 (near the Liyuan Subway Station) have the lowest rank. That is because all of them have a small population density. From Figure 8, we can see that spot 32 has the least changes. That is because the worst input factor is the condition of the pedestrian walkway, which is considered in the 7th place among the eight input factors.

From the abovementioned analysis, we find out that adding weight restrictions into the DEA framework is an effective method to integrate the DM's preference information. The unacceptable value could be avoided. However, we could not

Figure 7: Twelve current bike-sharing parking spots near the Liyuan Subway Station.

Table 3: The feature of these three places.

	District	Riding demand	Surrounding
Zhongguancun Science Park	Haidian district	Large	Commercial buildings, shopping mall, hospital, and university
Guomao CBD	Chaoyang district	Medium	Shopping mall, residential area, park, and commercial buildings
Liyuan Subway Station	Tongzhou district	Small	Residential area, university, hospital, and shopping mall

Table 4: The weights of different input factors.

Input factors	Weights	Ranks
Distance from the destination entrance	0.1338	4
Condition of the pedestrian walkway	0.0460	7
Proximity to the bus stations	0.2175	1
Proximity to the subway stations	0.2175	1
Proximity to the parking spots	0.1102	5
Distance from the bike lane	0.0396	8
The amount of available space	0.0649	6
Population density	0.1706	3

Table 5: The efficiency scores of spots 1 to12.

DMU	Basic BCC scores	Ranks	Weight-restricted DEA scores	Ranks
Spot 1	1	1	0.9396	9
Spot 2	1	1	1	1
Spot 3	1	1	1	1
Spot 4	0.6894	12	0.6773	12
Spot 5	1	1	1	1
Spot 6	1	1	1	1
Spot 7	1	1	1	1
Spot 8	1	1	1	1
Spot 9	1	1	1	1
Spot 10	1	1	0.9789	8
Spot 11	0.9022	10	0.8449	10
Spot 12	0.7460	11	0.6925	11
Average	0.9448		0.9295	

Table 6: The efficiency scores of spots 13 to 24.

DMU	Basic BCC scores	Ranks	Weight-restricted DEA scores	Ranks
Spot 13	0.8496	11	0.5706	11
Spot 14	0.9958	8	0.7660	10
Spot 15	1	1	1	1
Spot 16	1	1	1	1
Spot 17	1	1	1	1
Spot 18	0.9734	9	0.8982	9
Spot 19	1	1	1	1
Spot 20	0.5647	12	0.5017	12
Spot 21	1	1	0.9884	7
Spot 22	1	1	1	1
Spot 23	0.9632	10	0.9155	8
Spot 24	1	1	1	1
Average	0.9456		0.8867	

Table 7: The efficiency scores of spots 25 to 36.

DMU	Basic BCC scores	Ranks	Weight-restricted DEA scores	Ranks
Spot 25	1	1	0.9061	9
Spot 26	0.7736	12	0.6467	12
Spot 27	1	1	1	1
Spot 28	0.9407	10	0.8741	10
Spot 29	1	1	0.9872	6
Spot 30	1	1	1	1
Spot 31	1	1	0.9801	7
Spot 32	1	1	0.9934	5
Spot 33	1	1	1	1
Spot 34	1	1	0.9645	8
Spot 35	0.8895	11	0.8646	11
Spot 36	1	1	1	1
Average	0.9670		0.9347	

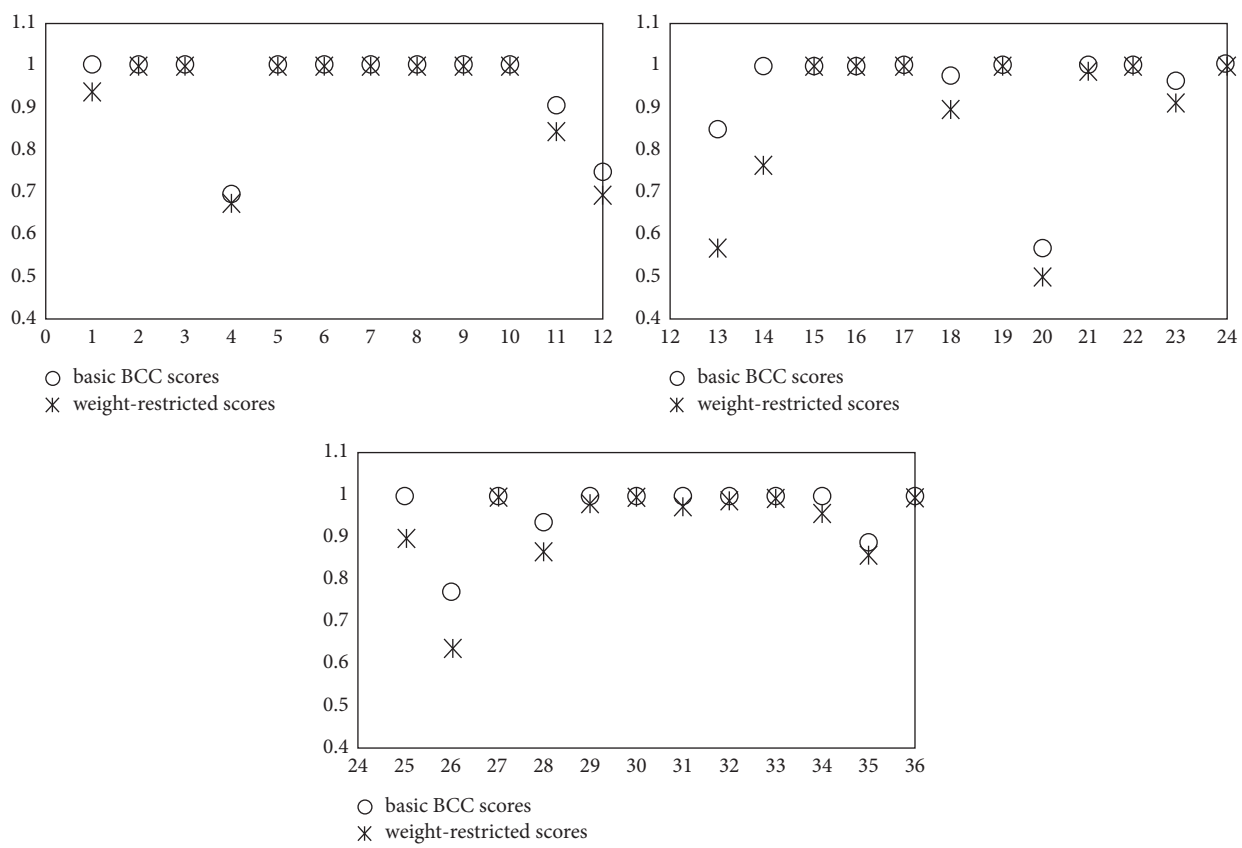


Figure 8: Result analysis.

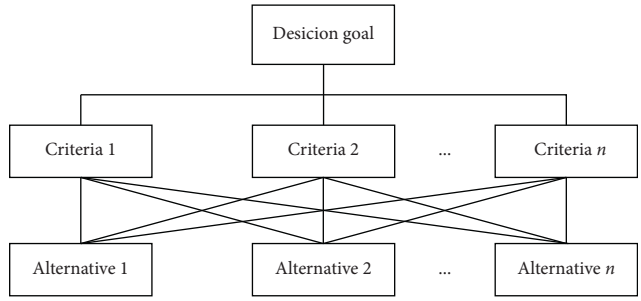


Figure 9: The hierarchy analysis tree.

rank the points with an efficient score of 1 in this model. In future work, we could utilize the method such as super DEA to overcome the deficiencies.

4.4. The Managerial Suggestions. For the final design, we suggest that the municipal planners can select the spots with high scores to build the virtual sites in the certain regions.

Nearing Zhongguancun Science Park, there are seven spots, spot 2, 3, and spot 6 to 10 on the efficiency frontier by incorporating the DMA' preference. It means that these spots are more efficient than the others. Besides, nine spots have the larger scores than the average efficient scores. This implies that the Beijing municipal could choose these spots to setup the electrical fences. In the region of the Guomao CBD, spot 15, 16, 17, 19, 22, and 24 are the efficient spots. Eight spots have the higher scores than the average. Across the Liyuan Subway Station, there are four alternative spots with the efficient score of 1 and eight spots above the average level of the twelve alternatives. All of these results have considered the multiple factors as we 2have mentioned in Section 3.

5. Conclusions

Bike-sharing parking points evaluation and selection have been identified as an important problem that can affect the efficiency of company operation and social warfare. It can be seen that selecting a parking space is a complicated problem, and DMs must have a wide view concerning qualitative and quantitative features to assess the symmetrical impact of the criteria to reach the most accurate result.

In this paper, four main criteria and eight subcriteria were determined by extensive literature examination and the insights of experts. The AHP was applied to obtain the weights of the criteria, as expected, and the proximity to the public traffic network and the population density were determined to be more important than other criteria. Besides, the AHP technique was also used to obtain the quality input factors. Then, the weight-restricted DEA using the AR I weighting scheme, which obtained from the AHP result, was proposed to evaluate the current bike-sharing parking spots in the final stage. Finally, 36 current parking spots near Zhongguancun Science Park, the Guomao CBD, and the Liyuan Subway Station were evaluated by using the combined method. From the results, we can find that not all the current parking sites are efficient in the performance. When DMs give a different performance in the evaluation factors, the alternatives will get different performance scores. This research also presents useful guidelines for layout optimization processes in other situations.

The hybrid method proposed in this paper cannot effectively rank the DMU if the efficiency value is 1. For example, there are seven spots at the rank of 1 near Zhongguancun Science Park. In future work, it would be interesting to adopt the superefficiency DEA or other methods to tackle this problem.

Appendix

(A). AHP Method

The process of applying the AHP is divided into the following four steps:

(i) *Step 1.* Constructing the hierarchy analysis trees

The hierarchy analysis tree is shown in Figure 9. There are three layers such as the target layer, the criteria layer, and the alternative layer.

(ii) *Step 2.* Building the pairwise comparison judgment matrices

The pairwise comparison matrix B, in which the element b_{ij} of the matrix is the relative importance of the i_{th} factor with respect to the j_{th} factor, is constructed by DMs and experts according to the scale given in Table 8.

$$B = \begin{bmatrix} 1 & b_{12} & \cdots & b_{1n} \\ 1/b_{12} & 1 & \cdots & b_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ 1/b_{1n} & 1/b_{2n} & \cdots & 1 \end{bmatrix}. \tag{A.1}$$

(iii) *Step 3.* Determining the weights of decision elements

The eigenvector method is used to yield priorities for criteria and for alternatives by criteria.

Table 8: Pairwise comparison scale.

Scale of importance	Interpretation
1	Equal importance
3	Moderate importance
5	Strong importance
7	Very strong or demonstrated importance
9	Extreme importance
2, 4, 6, 8	For compromise between the abovementioned values

Table 9: Average random consistency random index.

n	1	2	3	4	5	6	7	8	9	10
Random consistency	0	0	0.58	0.9	1.12	1.24	1.32	1.41	1.45	1.49

$$BW = \lambda_{\max} W, \tag{A.2}$$

where $\lambda_{\max}$ is the largest eigenvector and W is the weight vector. Note that we need to normalize W after calculating.

(iv) *Step 4*. Checking the consistency

We need to calculate the Consistency Ratio (CR) of each used pairwise matrix to make sure their values are less than 0.1.

$$\mathrm{CR} = \frac{\mathrm{CI}}{\mathrm{RI}}. \tag{A.3}$$

Consistency Index (CI) is the consistency of the judgment matrix and is defined as follows:

$$\mathrm{CI} = \frac{\lambda_{\max} - n}{n - 1}, \tag{A.4}$$

where n is the size of the matrix and Random Index (RI) is the average random consistency random index, as shown in Table 9.

(B). Basic BCC Model

Data envelopment analysis (DEA) is the linear programming (LP) technique that converts multioutput and multi-input into a scale measure of efficiency. It is suitable for the case of evaluating the current FF-BBS parking spots. The DEA model was originally proposed by Charnes et al. [54], which assumes a constant return to scales (CRS) called the CCR model. Banker et al. [55] extended the CCR model by adding the convexity constraint $1^T\lambda = 1$ to account for variable returns to scale (VRS), called the BCC model.

The multiplier form of the input-oriented BCC model is as follows. Assume that there are s DMUs, with each DMU producing m outputs by using n inputs. Y is the $m \times s$ output matrix, and X is the $n \times s$ input matrix. Y_0 and X_0 are the vectors of the DMU being evaluated. The $m \times 1$ vector u and the $n \times 1$ vector v are the flexible weights for outputs and inputs separately. u_* is the correspondent dual variables of the convexity constraint $1^T\lambda = 1$ (where λ is an intensity vector, "1" on the left hand is a $s \times 1$ vector, and 1 on the right hand is a scalar). The symbol 1 in equation (B.2) is a scalar and is a $s \times 1$ vector in equation (B.3). The objective is to obtain the most favorable weights of v and u such that the efficiency measure for the measured DMU is maximized.

$$\max \quad Y_0^T u + u_*, \tag{B.1}$$

$$\text{s.t.} \quad X_0^T v = 1, \tag{B.2}$$

$$-X^T v + Y^T u + 1u_* \le 0, \tag{B.3}$$

$$v, u \ge 0, \quad u_* \text{ free}. \tag{B.4}$$

Data Availability

The data used in this research are available upon request to the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (Grant no. 71801013), the Beijing Social Science Foundation (Grant no. 18GLC078), and the Fundamental Research Funds for the Central Universities (Grant no. 2020JBW004).

References

[1] Y. Liu, W. Y. Szeto, and S. C. Ho, "A static free-floating bike repositioning problem with multiple heterogeneous vehicles, multiple depots, and multiple visits," *Transportation Research Part C: Emerging Technologies*, vol. 92, pp. 208–242, 2018.

[2] M. Kabak, M. Erbaş, C. Çetinkaya, and E. Özceylan, "A GIS-based MCDM approach for the evaluation of bike-share stations," *Journal of Cleaner Production*, vol. 201, pp. 49–60, 2018.

[3] Y. Ma, J. Lan, T. Thornton, D. Mangalagiu, and D. Zhu, "Challenges of collaborative governance in the sharing economy: the case of free-floating bike sharing in Shanghai," *Journal of Cleaner Production*, vol. 197, pp. 356–365, 2018.

[4] A. Pal and Y. Zhang, "Free-floating bike sharing: solving real-life large-scale static rebalancing problems," *Transportation Research Part C: Emerging Technologies*, vol. 80, pp. 92–116, 2017.

[5] J. Liu, L. Qiao, Q. Meng et al., "Station site optimization in bike sharing systems," in *Proceedings of the 2015 IEEE*

International Conference on Data Mining, Atlantic City, NJ, USA, November 2015.

[6] Z. Sun, Y. Li, and Y. Zuo, "Optimizing the location of virtual stations in free-floating bike-sharing systems with the user demand during morning and evening rush hours," *Journal of Advanced Transportation*, vol. 2019, Article ID 4308509, 11 pages, 2019.

[7] C.-C. Hsu, J. J. H. Liou, H.-W. Lo, and Y.-C. Wang, "Using a hybrid method for evaluating and improving the service quality of public bike-sharing systems," *Journal of Cleaner Production*, vol. 202, pp. 1131–1144, 2018.

[8] K. Frenken and J. Schor, "Putting the sharing economy into perspective," *Environmental Innovation and Societal Transitions*, vol. 23, pp. 3–10, 2017.

[9] I. Frade and A. Ribeiro, "Bike-sharing stations: a maximal covering location approach," *Transportation Research Part A: Policy and Practice*, vol. 82, pp. 216–227, 2015.

[10] J. C. Garcıa-Palomares, J. Gutierrez, and M. Latorre, "Optimizing the location of stations in bike-sharing programs: a GIS approach," *Applied Geography*, vol. 35, no. 1-2, pp. 235–246, 2012.

[11] S. Mete, Z. A. Cil, and E. Özceylan, "Location and coverage analysis of bike-sharing stations in university campus," *Business Systems Research Journal*, vol. 9, no. 2, pp. 80–95, 2018.

[12] A. Giovannini, F. Malucelli, and M. Nonato, "Cycle-tourist network design," *Transportation Research Procedia*, vol. 22, pp. 154–163, 2017.

[13] L. M. Martinez, L. Caetano, T. Eiró, and F. Cruz, "An optimisation algorithm to establish the location of stations of a mixed fleet biking system: an application to the city of lisbon," *Procedia—Social and Behavioral Sciences*, vol. 54, pp. 513–524, 2012.

[14] J. G. Jin, H. Nieto, and L. Lu, "Robust bike-sharing stations allocation and path network design: a two-stage stochastic programming model," *Transportation Letters*, pp. 1–10. In press, 2019.

[15] M. Merakli and H. Yaman, "Robust intermodal hub location under polyhedral demand uncertainty," *Transportation Research Part B: Methodological*, vol. 86, pp. 66–85, 2016.

[16] P. Vogel, T. Greiser, and D. C. Mattfeld, "Understanding bike-sharing systems using data mining: exploring activity patterns," *Procedia—Social and Behavioral Sciences*, vol. 20, pp. 514–523, 2011.

[17] T. Wuerzer, S. Mason, and R. Youngerman, "Boise bike share location analysis," Boise State University, Boise, ID, USA, 2012.

[18] N. Mohajeri and G. R. Amin, "Railway station site selection using analytical hierarchy process and data envelopment analysis," *Computers & Industrial Engineering*, vol. 59, no. 1, pp. 107–114, 2010.

[19] M. H. M. Javadi, M. Ghandehari, and V. Hamidi Pouyandeh, "Locating of bicycle stations in the city of Isfahan using mathematical programming and multi-criteria decision making techniques," *International Journal of Academic Research in Accounting, Finance and Management Sciences*, vol. 3, no. 3, pp. 18–26, 2013.

[20] T. Kanjanakorn and M. Piantanakulchai, "Prioritizing suitable locations of bike sharing station by using the analytic hierarchy process (AHP)," in *Proceedings of the International Symposium on the Analytic Hierarchy Process*, Kuala Lumpur, Malaysia, June 2013.

[21] A. S. Loron and M. S. Loron, "An integrated fuzzy analytic hierarchy process-fuzzy data envelopment analysis (FAHP-FDEA) method for intelligent building assessment," *Tehnicki Vjesnik—Technical Gazette*, vol. 22, no. 2, pp. 383–389, 2015.

[22] S. H. Mousavi-Nasab and A. Sotoudeh-Anvari, "A comprehensive MCDM-based approach using TOPSIS, COPRAS and DEA as an auxiliary tool for material selection problems," *Materials & Design*, vol. 121, pp. 237–253, 2017.

[23] E. Ali Askari and M. Bashiri, "Design of a public bicycle-sharing system with safety," *Computational and Applied Mathematics*, vol. 36, no. 2, pp. 1023–1041, 2017.

[24] N. Bozorgi and M. Abedzadeh, "A multiple criteria facility layout problem using data envelopment analysis," *Management Science Letters*, vol. 1, pp. 363–370, 2011.

[25] W. M. Bowen, "Subjective judgements and data envelopment analysis in site selection," *Computers, Environment and Urban Systems*, vol. 14, no. 2, pp. 133–144, 1990.

[26] T. L. Saaty, *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*, McGraw-Hill, New York, NY, USA, 1980.

[27] T. Yang and C. Kuo, "A hierarchical AHP/DEA methodology for the facilities layout design problem," *European Journal of Operational Research*, vol. 147, no. 1, pp. 128–136, 2003.

[28] J. Shang and T. Sueyoshi, "A unified framework for the selection of a flexible manufacturing system," *European Journal of Operational Research*, vol. 85, no. 2, pp. 297–315, 1995.

[29] C. N. Wang, H. T. Tsai, T. P. Ho, V. T. Nguyen, and Y. F. Huang, "Multi-criteria decision making (MCDM) model for supplier evaluation and selection for oil production projects in Vietnam," *Processes*, vol. 8, no. 2, pp. 1–13, 2020.

[30] J. Korpela, A. Lehmusvaara, and J. Nisonen, "Warehouse operator selection by combining AHP and DEA methodologies," *International Journal of Production Economics*, vol. 108, no. 1-2, pp. 135–142, 2007.

[31] M. R. Khadivi and S. M. T. Fatemi Ghomi, "Solid waste facilities location using of analytical network process and data envelopment analysis approaches," *Waste Management*, vol. 32, no. 6, pp. 1258–1265, 2012.

[32] Y. Takamura and K. Tone, "A comparative site evaluation study for relocating Japanese government agencies out of Tokyo," *Socio-Economic Planning Sciences*, vol. 37, no. 2, pp. 85–102, 2003.

[33] R. Allen, A. Athanassopoulos, R. G. Dyson, and E. Thanassoulis, "Weights restrictions and value judgements in data envelopment analysis: evolution, development and future directions," *Annals of Operations Research*, vol. 73, pp. 13–34, 1997.

[34] M. P. L. Lins, A. C. Moreira da Silva, and C. A. K. Lovell, "Avoiding infeasibility in DEA models with weight restrictions," *European Journal of Operational Research*, vol. 181, no. 2, pp. 956–966, 2007.

[35] E. D. Mecit and I. Alp, "A new proposed model of restricted data envelopment analysis by correlation coefficients," *Applied Mathematical Modelling*, vol. 37, no. 5, pp. 3407–3425, 2013.

[36] V. V. Podinovski and A. D. Athanassopoulos, "Assessing the relative efficiency of decision making units using dea models with weight restrictions," *The Journal of the Operational Research Society*, vol. 49, no. 5, pp. 500–508, 1998.

[37] E. Murphy and J. Usher, "The role of bicycle-sharing in the city: analysis of the Irish experience," *International Journal of Sustainable Transportation*, vol. 9, no. 2, pp. 116–125, 2015.

[38] C. Cetinkaya, "Bike sharing station site selection for gaziantep," *Sigma Journal of Engineer & Nature Science*, vol. 35, no. 3, pp. 535–543, 2017.

[39] N. Subramanian and R. Ramanathan, "A review of applications of analytic hierarchy process in operations management," *International Journal of Production Economics*, vol. 138, no. 2, pp. 215–241, 2012.

[40] R. Stawicki and K. D. Lawrence, "Multiple criteria dea with and without weighting restrictions," *Applications in Multi-criteria Decision Making, Data Envelopment Analysis, and Finance*, Emerald Publishing Limited, Bingley, UK, pp. 95–104, 2010.

[41] T. Joro and P. J. Korhonen, *Extension of Data Envelopment Analysis With Preference Information: Value Efficiency*, Springer, Boston, MA, USA, 2015.

[42] W. Cook, A. Kazako, Y. Roll, and L. Seiford, "A data envelopment approach to measuring efficiency: case analysis of highway maintenance patrols," *The Journal of Socio-Economics*, vol. 20, no. 1, pp. 83–103, 1991.

[43] P. Zhou, B. Ang, and K. Poh, "Measuring environmental performance under different environmental DEA technologies," *Energy Economics*, vol. 30, no. 1, pp. 1–14, 2008.

[44] C. Kao and H. Hung, "Efficiency analysis of university departments: an empirical study☆," *Omega*, vol. 36, no. 4, pp. 653–664, 2008.

[45] J. Soleimani-Damaneh, M. Hamidi, and H. Sajadi, "Evaluating the performance of Iranian football teams utilizing linear programming," *American Journal of Operations Research*, vol. 1, no. 2, pp. 65–72, 2011.

[46] J. E. Beasley, "Comparing university departments," *Omega*, vol. 18, no. 2, pp. 171–183, 1990.

[47] Y. Roll, W. D. Cook, and B. Golany, "Controlling factor weights in data envelopment analysis," *IIE Transactions*, vol. 23, no. 1, pp. 2–9, 1991.

[48] V. V. Podinovski, "Side effects of absolute weight bounds in DEA models," *European Journal of Operational Research*, vol. 115, no. 3, pp. 583–595, 1999.

[49] Y.-H. B. Wong and J. E. Beasley, "Restricting weight flexibility in data envelopment analysis," *The Journal of the Operational Research Society*, vol. 41, no. 9, pp. 829–835, 1990.

[50] R. G. Thompson, L. N. Langemeier, C. T. Lee, E. Lee, and R. M. Thrall, "The role of multiplier bounds in efficiency analysis with application to Kansas farming," *Journal of Econometrics*, vol. 46, no. 1-2, pp. 93–108, 1990.

[51] G. E. Halkos, N. G. Tzeremes, and S. A. Kourtzidis, "Weight assurance region in two-stage additive efficiency decomposition DEA model: an application to school data," *Journal of the Operational Research Society*, vol. 66, no. 4, pp. 696–704, 2015.

[52] R. G. Thompson, F. D. Singleton, R. M. Thrall, and B. A. Smith, "Comparative site evaluations for locating a high-energy physics lab in Texas," *Interfaces*, vol. 16, no. 6, pp. 35–49, 1986.

[53] D. Chang and J. Y. Wang, "A new model to improve the subjectivity of AHP," *Value Engineering*, vol. 9, pp. 32–34, 2004.

[54] A. Charnes, W. W. Cooper, and E. Rhodes, "Measuring the efficiency of decision making units," *European Journal of Operational Research*, vol. 2, no. 6, pp. 429–444, 1978.

[55] R. D. Banker, A. Charnes, and W. W. Cooper, "Some models for estimating technical and scale inefficiencies in data envelopment analysis," *Management Science*, vol. 30, no. 9, pp. 1078–1092, 1984.

Hindawi
Discrete Dynamics in Nature and Society
Volume 2020, Article ID 3151949, 13 pages
https://doi.org/10.1155/2020/3151949

Research Article

Evaluating the Sustainable Growth of Small- and Medium-Sized Construction Enterprises Using the Multicriteria Decision-Making Method

Wenbao Wang,[1] Wenhe Lin ,[2] Xinyi Dai,[2] Yingzheng Yan,[2] and Aiping Xuan[2]

[1]*College of Civil Engineering, Yango University, Fuzhou, China*
[2]*College of Management & College of Tourism, Fujian Agriculture and Forestry University, Fuzhou, China*

Correspondence should be addressed to Wenhe Lin; 253062775@qq.com

Received 23 June 2020; Revised 20 August 2020; Accepted 26 August 2020; Published 9 September 2020

Academic Editor: Fuqiang Gu

Sustainable growth plays an important role in the development of small- and medium-sized construction enterprises and has increasingly become a hot topic in the political and academic arenas. Growth is an important indicator for measuring the operational status and development potential of listed enterprises and is of great significance for studying sustainable growth. Considering the development of small- and medium-sized construction enterprises, a growth evaluation index system that includes growth resources, growth ability, and growth innovation was established, and a growth evaluation model was constructed using the improved entropy-VIKOR algorithm. Evaluation and analysis of the sustainable growth of small- and medium-sized construction enterprises listed before the end of 2012 were conducted. The parameter sensitivity, weight sensitivity, and effectiveness of the evaluation model were analysed. The empirical results showed that compared with the TOPSIS method and the traditional VIKOR method, the multicriteria evaluation model for the sustainable growth of small- and medium-sized construction enterprises proposed in this paper had better stability and effectiveness. Finally, the key influencing factors, growth ranking, and dynamic growth types of small- and medium-sized construction enterprises were analysed, and targeted suggestions were proposed. The research results can provide a reference for managers, investors, and regulators.

1. Introduction

Small- and medium-sized enterprises (SMEs) constitute an important part of the overall national economy, and the growth of SMEs is the internal impetus for the development of enterprises and the national economy. The growth of SMEs is an important indicator used to measure the operating conditions and development potential of listed enterprises. At present, small- and medium-sized construction enterprises account for more than 90% of the total number of construction enterprises in China [1], and they have become an indispensable part of China's construction industry. However, small- and medium-sized construction enterprises lag far behind large construction enterprises in terms of the overall professional level, management systems, scale effects, and technological innovation ability. Therefore, the sustainable growth of small- and medium-sized construction enterprises is becoming a hot research topic and is related not only to the interests of managers and investors but also to the economic development of the entire country. With the development of the capital market, the evaluation of the sustainable growth of small- and medium-sized construction enterprises has become increasingly complicated and a key focus for firms, investors, and government management departments.

Enterprise growth refers to the business capability, sustainable development ability, and, importantly, the evaluation of the future development of an enterprise. In terms of dynamic potential, enterprise growth can be divided into sustainable growth and interrupted growth (maintenance and reduction). Sustainable growth is reflected in the continuous expansion of the market scale of the industry and

the expansion of the enterprise's market share and scale of business; the increase in operating benefits year over year; the continuous improvement of the enterprise's system, culture, and organization; the enterprise's adaptability to the internal and external environments. Many scholars around the world have examined the growth of enterprises and revealed important findings. In China, Wang [2] used China's listed companies in the environmental protection industry as an example and selected 14 financial indicators, including profitability, financial risk, operating capacity, and growth level, for principal component analysis, arguing that growth is greatly affected by profitability. Hu et al. [3] studied enterprise growth using the principal component projection method based on multi-index comprehensive evaluation. Gong and Lv [4] proposed a model for enterprise growth evaluation based on R language data analysis. Xia et al. [5–10] studied the growth of SMEs based on factor analysis. Zhao et al. [11] established an index system based on the dynamic ability angle and used the fuzzy comprehensive evaluation method to construct an evaluation model of the growth of high-tech enterprises. In addition, Cao and Chang analysed the influencing factors of enterprise growth [12, 13]. In other countries, Zeng [14] studied the evaluation of enterprises' technological innovation capabilities using two-tuple linguistic information. Liu et al. [15] established a growth quality model for LED listed companies on the basis of industrial innovation theory, corporate governance theory, and enterprise organization theory. Using the network analysis method, Lahdelma and Laakso [16] conducted a network evaluation of companies. Additionally, Lefley and Sarkis [17] applied the FAP model to the evaluation of strategic information technology projects. Thus, the literature focuses mainly on the empirical analysis of the factors related to the growth of enterprises and seldom examines the construction of an evaluation model. Many studies consider the evaluation of enterprise growth in China through the construction of index systems and evaluation models. However, most growth evaluation research adopts a single evaluation method (e.g., analytic hierarchy process, principal component analysis, mutation progression method, and ideal point method), and a method is selected based on the application of mathematics, without identifying the validity of the method itself or the reference value of the investigation results. Therefore, it is necessary to use the combination evaluation method to study the evaluation of enterprises' sustainable growth.

2. Research Methods and Evaluation Indicators

2.1. Research Methods. Compared to the traditional single evaluation method, the multicriteria decision-making compromise method has the advantages of intuitive analysis principles and simple calculations [18]. Additionally, entropy weight and multicriteria decision-making evaluation methods, especially the entropy-TOPSIS method, have been widely used in research on urban industry competitiveness, regional system development, and even enterprise evaluation. Moreover, the VIKOR method outperforms the TOPSIS method in multicriteria decision-making because it can rank alternatives according to their proximity to the ideal solution, while the TOPSIS method cannot account for the importance of alternative solutions or their proximity to the positive and negative ideal solutions. Unlike the TOPSIS method, the VIKOR method also considers the maximization of group utility and the minimization of individual regrets simultaneously, as well as the subjective preferences of decision makers, and it provides a more reasonable solution that can achieve a compromise among priorities when dealing with ranking problems [19]. In recent years, the VIKOR method has been applied in several fields, such as sustainable and renewable energy [20], supplier selection [21], and project selection [22]. It is also combined with attribute weight determination methods. For example, the analytic hierarchy process can be combined with the VIKOR eclectic ranking method to construct a flexible multicriteria decision framework [23]. Entropy weight and the VIKOR method are also combined for use in project risk assessment [24]. An entropy-VIKOR combination method has been proven able to solve multiattribute evaluation and decision problems. However, the traditional VIKOR method uses the information regarding only the positive ideal solution, not the negative ideal solution. Although the positive ideal solution and the negative ideal solution often have a certain degree of correlation, the negative ideal solution still has its own unique information. Therefore, we should make full use of the information on the negative ideal solution to ensure the rationality of the decision results. Therefore, in this paper, the VIKOR method is improved to make full use of positive and negative ideal solution information to evaluate the sustainable growth of small- and medium-sized construction enterprises, and the practicable research results can serve as a reference for managers, investors, and regulators.

The improved entropy-VIKOR algorithm adopted in this paper is constructed as follows:

(1) Establishment of the decision matrix for the evaluation criteria.

It is assumed that there are n evaluation indicators and m evaluation objects for multiattribute decision-making problems. Matrix X is composed of evaluation index values, X_{ij} represents the jth index value of the ith evaluation object, and the decision matrix is as follows:

$$\left(X_{ij}\right)_{m\times n} = \begin{pmatrix} X_{11} & X_{21} & K & X_{m1} \\ X_{12} & X_{22} & K & X_{m2} \\ K & K & K & K \\ X_{1n} & X_{2n} & K & X_{mn} \end{pmatrix}. \tag{1}$$

(2) Standardization of the indicators.

Considering that the data for each indicator are of different dimensions, it is necessary to perform dimensionless processing on all indicator data. When the evaluation index has the criterion of benefit type (the larger the value, the better), x_{ij} is calculated as follows:

$$x_{ij} = \frac{x'_{ij} - \min_i\left(x'_{ij}\right)}{\max_i\left(x'_{ij}\right) - \min_i\left(x'_{ij}\right)}. \tag{2}$$

When the evaluation indicator has cost-type criteria (the smaller the value, the better), x_{ij} is calculated as follows:

$$x_{ij} = \frac{\max_i\left(x'_{ij}\right) - x'_{ij}}{\max_i\left(x'_{ij}\right) - \min_i\left(x'_{ij}\right)}. \tag{3}$$

Additionally, when the indicator has a fixed criterion (the closer the value is to an ideal value β, the better), x_{ij} is calculated as follows:

$$x_{ij} = \frac{1.0 - \left|x'_{ij} - \beta\right|}{\max_i\left|x'_{ij} - \beta\right|}. \tag{4}$$

(3) Calculating the normalized decision matrix.

The normalized decision matrix is expressed as follows:

$F = [p_{ij}]_{m\times n}$; in the formula,

$$p_{ij} = \frac{x_{ij}}{\sum_{i=1}^{m} x_{ij}}, \quad i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, n. \tag{5}$$

(4) Determining the entropy weight of the evaluation index.

First, the entropy value of index j is calculated:

$$E_j = -K \sum_{i=1}^{m} p_{ij} \ln p_{ij}, \quad i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, n. \tag{6}$$

In the formula, $k = (1/\ln m) > 0$, $0 \le E_j \le 1$. Then, the difference coefficient of the jth indicator is calculated: $g_j = 1 - E_j$. If j is to keep a certain value, the smaller the difference g_j of X_{ij} is, the larger the value of E_j is. From this, the weight of each indicator can be defined as follows:

$$W_j = \frac{g_j}{\sum_{j=1}^{n} g_j}. \tag{7}$$

(5) Determining the positive ideal solution and the negative ideal solution.

If the jth criterion has the property of the larger, the better, then the positive ideal solution set A^* and the negative ideal solution set A^- can be expressed as follows:

$$\begin{aligned} A^* &= \left\{\max_{j=1}^{n} f_{ij} \,\middle|\, i = 1, 2, \ldots, m\right\} \\ &= \left\{f_1^*, f_2^*, \ldots, f_j^*, \ldots, f_n^*\right\}, \\ A^- &= \left\{\min_{j=1}^{n} f_{ij} \,\middle|\, i = 1, 2, \ldots, m\right\} \\ &= \left\{f_1^-, f_2^-, \ldots, f_j^-, \ldots, f_n^-\right\}. \end{aligned} \tag{8}$$

Additionally, if smaller is better for the jth criterion, then the positive ideal solution set A^* and the negative ideal solution set A^- can be expressed as follows:

$$\begin{aligned} A^* &= \left\{\min_{j=1}^{n} f_{ij} \,\middle|\, i = 1, 2, \ldots, m\right\} \\ &= \left\{f_1^*, f_2^*, \ldots, f_j^*, \ldots, f_n^*\right\}, \\ A^- &= \left\{\max_{j=1}^{n} f_{ij} \,\middle|\, i = 1, 2, \ldots, m\right\} \\ &= \left\{f_1^-, f_2^-, \ldots, f_j^-, \ldots, f_n^-\right\}. \end{aligned} \tag{9}$$

(6) The group utility value and the individual regret value are then calculated. With the positive ideal solution as a reference, the ratio of the distance between the ith scheme and the positive ideal solution is as follows:

$$S_i^+ = \sum_{j=1}^{n} w_j \frac{\left(f_j^* - f_{ij}\right)}{\left(f_j^* - f_j^-\right)}. \tag{10}$$

The ratio of the distance between the ith scheme and the negative ideal solution is as follows:

$$R_i^+ = \max_j \left[w_j \frac{\left(f_j^* - f_{ij}\right)}{\left(f_j^* - f_j^-\right)}\right]. \tag{11}$$

Taking the negative ideal solution as a reference, the distance ratio between the ith scheme and the positive ideal solution is as follows:

$$S_i^- = \sum_{j=1}^{n} w_j \frac{\left(f_{ij} - f_j^-\right)}{\left(f_j^* - f_j^-\right)}. \tag{12}$$

The distance ratio between the ith scheme and the negative ideal solution is as follows:

$$R_i^- = \max_j \left[w_j \frac{\left(f_{ij} - f_j^-\right)}{\left(f_j^* - f_j^-\right)}\right]. \tag{13}$$

The evaluation values of group utility and individual regret are as follows:

$$\begin{aligned} S_i &= \frac{S_i^+}{S_i^-}, \\ R_i &= \frac{R_i^+}{R_i^-}, \end{aligned} \tag{14}$$

where w_j is the weight of the jth criterion, S_i can be seen as a consent decision indicator, and R_i can be seen as an opposition decision indicator. Assuming that every decision-maker can vote for or against every plan, the smaller the S_i scheme is, the more people agree with it, and the smaller the R_i scheme is, the fewer people oppose it. Therefore, smaller is better for S_i and R_i.

(7) Calculating the VIKOR comprehensive indicator Q and sorting the pros and cons.

The VIKOR comprehensive indicator of the ith scheme can be obtained by the following formula:

$$Q_i = v\left[\frac{S_i - S^*}{S^- - S^*}\right] + (1-v)\left[\frac{R_i - R^*}{R^- - R^*}\right], \quad (15)$$

where $S^* = \min_i S_i$, $S^- = \max_i S_i$, $R^* = \min_i R_i$, $R^- = \max_i R_i$, and v represents the weight of the distance ratio between the ideal solution and the negative ideal solution (the value is usually set to 0.5). In the above equation, v represents the decision-making mechanism coefficient. When $v > 0.5$, it means that the decision is made according to the majority of resolutions. When $v = 0.5$, it means that the decision is made according to agreement. When $v < 0.5$, it means that the decision is made in favour of the opposition. In VIKOR, we set v to be 0.5 when pursuing the maximization of group utility and the minimization of individual regret.

(8) According to the VIKOR method, the selected objects are sorted according to S, R, and Q values, from small to large.

(9) Acquiring the best ranking of selection objects. According to the order of Q from small to large, the first selection object X^1 is the best selection object when it meets the following two conditions:

Condition 1 (acceptance advantage): $\mathrm{DQ} = (1/(n-1))$ meets $Q(X^2) - Q(X^1) \geq \mathrm{DQ}$.

Condition 2 (stability of decision): the best selection object X^1 must be the best according to the Q value ranking and the S or R value ranking at the same time to ensure the stability of the decision.

If the above conditions cannot be met at the same time, there is a compromise satisfactory solution set (that is, there is more than one satisfactory solution), which can be divided into two situations:

(1) If condition 1 is not met, $X^1, X^2, \ldots, X^n$ constitute a set of alternative compromise satisfactory solutions. X^n is the maximum n value determined by $Q(X^n) - Q(X^1) < \mathrm{DQ}$.

(2) If condition 2 is not satisfied, then X^1 and X^2 constitute a set of alternative compromise satisfactory solutions; that is, X^1 and X^2 can be selected as satisfactory solutions.

2.2. Construction of the Evaluation Index System. Penrose defined an enterprise as a resource aggregate that relies on a particular management structure and whose economic benefits are based on resources and capabilities; thus, we establish an analysis framework of enterprise resources, enterprise capabilities, and enterprise growth [19]. Hall divided enterprise resources into tangible resources, intangible resources, and capabilities [24]. Enterprise capability itself is also a system of rich content, and Lavie described the process of capability reconstruction as a two-stage process model based on the zero-order capability level and considering the capability gap [25]. Based on Penrose's theory of enterprise growth, this paper establishes the growth evaluation index system for small- and medium-sized construction enterprises from the three aspects of growth resources, growth capability, and growth innovation (as shown in Table 1) by referring to the research results of scholars in recent years [26–28] and considering the characteristics of listed construction enterprises, such as economies of scale, profitability, labour intensity, resource consumption, sensitivity to policy and the ecological environment, profitability, operation, concomitant development and expansion, consistency of debt payment, and risk-averse behaviour [29]. The indicator system consists of 3 criteria layers, 10 factor layers, and a total of 25 indicators.

2.3. Types of Dynamic Growth of Enterprises. To explore the growth process of small- and medium-sized construction enterprises horizontally, it is necessary to extend the time span considered. The growth of an enterprise can be dynamically divided into continuous growth and interrupted growth (maintenance and reduction) [30]. This paper proposes a feasible classification method for reference. The growth rankings of small- and medium-sized construction enterprises from 2013 to 2016 are denoted as X_1, X_2, X_3, and X_4 and are calculated using the annual growth rate to obtain the increment (or change amount) of the new ranking of enterprises, $I_i = -((X_{i+1} - X_i)/X_i)$ (i = 1, 2, 3, 4).

Based on the annual growth rate, the mean and variance (SD) of the enterprises' annual growth rate are calculated. The mean and variance of the growth rate are the main basis for the vertical comparison of the growth of small- and medium-sized construction enterprises in this paper. The calculation formulas are as follows:

$$\mathrm{mean} = \sqrt[3]{(1+I_1)(1+I_2)(1+I_2)} - 1,$$
$$\mathrm{SD} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(I_i - \overline{I})^2}. \quad (16)$$

Finally, the enterprises studied in this paper can be classified into six growth categories, including steady growth, fluctuating growth, steady negative growth, fluctuating negative growth, steady maintenance, and fluctuating maintenance. The basis for classification is as follows:

(A) Mean > 0, 且SD < average (SD) steady growth

(B) Mean > 0, 且SD > average (SD) fluctuating growth

(C) Mean < 0, 且SD < average (SD) steady negative growth

(D) Mean < 0, 且SD > average (SD) fluctuating negative growth

(E) Mean = 0, 且SD < average (SD) steady maintenance

(F) Mean = 0, 且SD > average (SD) fluctuating maintenance

Table 1: The growth evaluation index system of small- and medium-sized construction enterprises.

Target layer	Criteria layer	Factor layer	Indicator layer
The growth evaluation index system of small- and medium-sized construction enterprises	Growth resources	Intangible resources	Ratio of intangible assets, growth rate of intangible assets
		Tangible resources	Increasing rate of fixed assets, cash ratio, inventory, accounts receivable
		Human resources	Number of employees, quality of employees
	Growth ability	Operation capacity	Accounts receivable turnover, current asset turnover, inventory turnover, turnover of fixed assets
		Profitability	Operating gross profit margin, return on equity, return on assets
		Debt-paying ability	Asset-liability ratio, current ratio, quick ratio
		Marketing ability	Net profit margin on sales, market share
		Market anticipation capability	Price earning (PE) ratio, price/book (PB) value ratio
	Growth innovation	Technological innovation	Growth rate of research and development, input intensity
		Service innovation	Diversity of products or services

2.4. Sample Selection and Data Source. How to maintain rapid growth and steady and sustainable development in an environment of fierce competition is the realistic and challenging problem faced by small- and medium-sized construction enterprises. To ensure a sufficient sample size and considering the reliability and integrity of data sources, this study selects small- and medium-sized construction enterprises listed before December 31, 2012, as the sample. Excluding the enterprises with incomplete data, we selected 26 enterprises and analysed sample data from four consecutive years: 2013, 2014, 2015, and 2016. The sample is shown in Table 2. The indicator data are obtained from the WIND database; these data include the information disclosed by the sample companies in their annual reports.

3. Empirical Research Results and Analysis

3.1. Evaluation Results for Enterprise Growth

3.1.1. The Sample Data Were Processed following the Steps of the Entropy-VIKOR Algorithm. The weight value of each index layer is calculated using formulas (6) and (7). The weight of the factor layer is the sum of the weights of the corresponding index layer, and the weight of the criterion layer is the sum of the weights of the corresponding factor layer. The weight values of the criterion layer, factor layer, and index layer are shown in Table 3.

Table 3 shows that the contribution rate of the operation capacity index is the largest, with a weight of 0.3443, indicating that enterprise growth needs to be judged by considering multiple aspects and cannot be determined by a single index. Additionally, intangible resources, tangible resources, and human resources accounted for a relatively large proportion, with weights of 0.2917, 0.1184, and 0.0864, respectively, indicating that these resources to a large extent represent the growth capacity of enterprises. This finding shows that these four factors have a strong influence on the growth of small- and medium-sized construction enterprises.

The positive ideal solution and the negative ideal solution of each index are determined, and the VIKOR comprehensive index Q is calculated. Taking 2016 as an example, $Q(X^2) - Q(X^1) = 0.0855 - 0.000 = 0.0855 \geq \mathrm{DQ} = 1/(n-1) = 1/25 = 0.04$, and X^1 is the evaluation object with the top S or R ranking, so it needs to be ranked only according to the Q value [31]. The ranking of Q values in each year is shown in Table 4.

3.1.2. Enterprise Growth Rankings. Table 4 shows that the enterprises with the highest growth in 2016 were YABAITE, Gold Mantis, HONGTAO, QDDFTT, and ZCT, and the enterprises with the lowest growth were BXLQ, YT-ECO, GHE, ZSSF, and PALM. According to the data for the four years studied, the top four companies are Gold Mantis, ZCT, HONGTAO, and Orient Landscape, and the bottom four companies are BXLQ, YT-ECO, ZSSF, and GHE.

According to Section 2.3, which deals with the classification of dynamic growth types for enterprises, there are 6 small- and medium-sized construction enterprises with steady growth, 16 with steady negative growth, and 4 with fluctuating negative growth, as shown in Table 4.

3.2. Effectiveness Analysis of Evaluation Methods. To verify the effectiveness of this method, it is compared with the TOPSIS method and the traditional VIKOR method in this paper. TOPSIS is one of the methods most widely used for multiattribute decision-making. By detecting the distance between the evaluation object and the optimal solution and the worst solution, the closeness degree is used to rank all schemes. Taking 2016 as an example, the ranking results according to each method are shown in Table 5.

Among them, the greater the relative closeness of the TOPSIS method is, the higher the ranking is; the smaller the Q value of the traditional VIKOR method and the improved VIKOR method is, the higher the ranking is. Table 5 shows that YABAITE ranks first and QDDFTT, Gold Mantis, and ZCT all rank in the top five, but the specific ranking

Table 2: The sample companies.

Number	Stock code	Company abbreviation	Number	Stock code	COHR
1	002047.SZ	BAUING	14	002323.SZ	YABAITE
2	002051.SZ	CAMC	15	002325.SZ	HONGTAO
3	002060.SZ	GHE	16	002375.SZ	YASHA
4	002061.SZ	ZCT	17	002431.SZ	PALM
5	002062.SZ	HR	18	002469.SZ	Sanwei
6	002081.SZ	Gold Mantis	19	002482.SZ	GRANDLAND
7	002116.SZ	China Haisum	20	002541.SZ	HOLU
8	002135.SZ	ZSSF	21	002542.SZ	ZHCGE
9	002140.SZ	DHC	22	002545.SZ	QDDFTT
10	002178.SZ	Yanhua Smartech	23	002586.SZ	RECLAM
11	002200.SZ	YT-ECO	24	002620.SZ	Ruihe
12	002307.SZ	BXLQ	25	002628.SZ	CDR&B
13	002310.SZ	Orient Landscape	26	002663.SZ	Pbland

Table 3: Index weight of each level.

Target layer	Criteria layer	Weight	Factor layer	Weight	Indicator layer	Weight
Growth index system for small- and medium-sized listed construction enterprises	Growth resources	0.4965	Intangible resources	0.2917	The proportion of intangible assets	0.0947
					Intangible asset growth rate	0.1970
					Growth rate of fixed assets	0.0494
			Tangible resources	0.1184	Cash ratio	0.0587
					Inventory	0.0060
					Accounts receivable	0.0043
			Human resources	0.0864	Number of employees	0.0624
					Staff quality	0.0240
	Growth ability	0.4562	Operation capacity	0.3443	Accounts receivable turnover rate	0.0851
					Current asset turnover	0.0346
					Inventory turnover	0.1589
					Fixed asset turnover	0.0657
			Profitability	0.0292	Gross operating margin	0.0127
					Return on equity	0.0049
					Return on assets	0.0116
			Debt-paying ability	0.0261	Asset-liability ratio	0.0169
					Current ratio	0.0050
					Quick ratio	0.0042
	Growth innovation	0.0474	Marketing capability	0.0520	Net profit margin	0.0095
					Market occupancy	0.0425
			Market expectation ability	0.0046	Price earnings ratio	0.0026
					Price/book value ratio	0.0020
			Technology innovation	0.0459	Growth rate of R&D cost	0.0123
					Intensity of R&D input	0.0336
			Service innovation	0.0015	Diversity of products or services	0.0015

differs by the method. Furthermore, from Table 6, it can be seen that the correlation coefficient of the three methods is high, and thus the evaluation results of the three methods are relevant and different, indicating the effectiveness and rationality of the improved method. The Q value of the improved VIKOR method is obviously smaller than that of the traditional VIKOR method. Taking only the positive ideal solution as the reference, the actual evaluation value is unreasonably high, indicating the importance of considering the information of the negative ideal solution to ensure rational decision-making; this figure also shows the superiority of the improved VIKOR method.

The purpose of the validity test is to verify the error between the evaluation result and the true value. Correlation

Table 4: Enterprise Q value ranking (2013–2016).

Enterprise	Years								Type of growth
	2013		2014		2015		2016		
	Q	Rank	Q	Rank	Q	Rank	Q	Rank	
BAUING	0.4981	11	0.3526	10	0.2250	6	0.3704	9	Steady negative growth
CAMC	0.7292	19	0.4391	15	0.5560	18	0.5628	13	Steady growth
GHE	0.5916	17	0.7736	23	0.7724	25	0.8860	24	Steady negative growth
ZCT	0.1439	4	0.0579	3	0.0839	4	0.1259	5	Steady negative growth
HR	0.5682	14	0.5177	17	0.5064	16	0.7175	21	Steady negative growth
Gold Mantis	0.0343	2	0.0145	2	0.0492	3	0.0855	2	Steady negative growth
China Haisum	0.4787	9	0.4303	14	0.3474	10	0.5541	12	Steady negative growth
ZSSF	0.8239	24	0.7044	22	0.6248	22	0.8012	23	Steady growth
DHC	0.5545	13	0.5517	19	0.5109	17	0.6127	18	Steady negative growth
Yanhua Smartech	0.5705	15	0.3927	12	0.0000	1	0.5465	11	Fluctuating negative growth
YT-ECO	0.7606	21	0.6917	20	0.6598	23	0.9002	25	Steady negative growth
BXLQ	0.8973	25	0.9196	25	1.0000	26	0.9996	26	Steady negative growth
Orient landscape	0.2362	6	0.1574	5	0.2143	5	0.2974	6	Steady negative growth
YABAITE	0.5114	12	0.3650	11	0.5674	19	0.0000	1	Steady negative growth
HONGTAO	0.1263	3	0.1775	6	0.0447	2	0.0901	3	Steady negative growth
YASHA	0.0000	1	0.4401	16	0.4167	12	0.5885	15	Fluctuating negative growth
PALM	0.7795	22	0.2314	8	0.6035	21	0.7841	22	Fluctuating negative growth
Sanwei	0.2904	7	0.3369	9	0.2379	7	0.3051	7	Steady negative growth
GRANDLAND	0.7541	20	0.6950	21	0.4703	14	0.5692	14	Steady growth
HOLU	0.5838	16	0.5179	18	0.5716	20	0.6219	19	Steady negative growth
ZHCGE	0.6413	18	0.0082	1	0.3613	11	0.5933	17	Fluctuating negative growth
QDDFTT	0.4925	10	0.4176	13	0.2801	8	0.1187	4	Steady growth
RECLAM	0.2025	5	0.1535	4	0.3153	9	0.4292	10	Steady negative growth
Ruihe	0.9205	26	0.8666	24	0.4602	13	0.3239	8	Steady growth
CDR&B	0.7892	23	1.0000	26	0.6780	24	0.5927	16	Steady growth
Pbland	0.3458	8	0.2220	7	0.4772	15	0.6751	20	Steady negative growth

Table 5: Ranking results of each method in 2016.

Enterprise	TOPSIS method		VIKOR method		Improved VIKOR method	
	Relative closeness C	Rank	Q	Rank	Q	Rank
BAUING	0.1166	13	0.8884	9	0.3704	9
CAMC	0.1070	15	0.9005	11	0.5628	13
GHE	0.0803	24	0.9825	23	0.8860	24
ZCT	0.2942	3	0.6911	4	0.1259	5
HR	0.1000	19	0.9816	22	0.7175	21
Gold Mantis	0.2909	5	0.6656	3	0.0855	2
China Haisum	0.1342	8	0.8889	10	0.5541	12
ZSSF	0.0927	22	0.9957	26	0.8012	23
DHC	0.1224	9	0.9510	19	0.6127	18
Yanhua Smartech	0.1192	10	0.9347	17	0.5465	11
YT-ECO	0.0772	26	0.9947	25	0.9002	25
BXLQ	0.0775	25	0.9826	24	0.9996	26
Orient Landscape	0.1661	7	0.7114	5	0.2974	6
YABAITE	0.5286	1	0.0000	1	0.0000	1
HONGTAO	0.2926	4	0.7170	6	0.0901	3
YASHA	0.1116	11	0.9307	16	0.5885	15
PALM	0.0970	21	0.9685	21	0.7841	22
Sanwei	0.1848	6	0.8807	8	0.3051	7
GRANDLAND	0.1040	17	0.9188	14	0.5692	14
HOLU	0.1067	16	0.9500	18	0.6219	19
ZHCGE	0.1084	14	0.9092	13	0.5933	17
QDDFTT	0.2954	2	0.4918	2	0.1187	4
RECLAM	0.1120	12	0.9252	15	0.4292	10
Ruihe	0.0974	20	0.8598	7	0.3239	8
CDR&B	0.0924	23	0.9644	20	0.5927	16
Pbland	0.1013	18	0.9075	12	0.6751	20

Table 6: The correlation between the 2016 evaluation results and the excess return rate ranking.

		TOPSIS method	Traditional VIKOR method	Improved VIKOR method
Excess return	Correlation coefficient	−0.151	0.143	0.120
	Sig.	0.460	0.485	0.559
	Number of enterprises	26	26	26
TOPSIS method	Correlation coefficient	1.000	0.832*	0.867**
	Sig.		0.000	0.000
	Number of enterprises	26	26	26
Traditional VIKOR method	Correlation coefficient		1.000	0.931**
	Sig.			0.000
	Number of enterprises		26	26

analysis of the annual excess return of enterprise stocks and the growth rankings of enterprises is one of the most commonly used methods in the literature. Based on the assumption of rational market investment, enterprise stocks with high growth rankings should be favoured by investors and have a high rate of return. The sample firms were ranked according to the annual excess rate of return, and the sequential data and evaluation results were analysed by Spearman correlation, as shown in Table 6. The results of the traditional VIKOR method and the improved VIKOR method are positively but nonsignificantly correlated with the ranking of the annual excess return of stocks, while the TOPSIS method is negatively but nonsignificantly correlated. This finding also shows that the evaluation results of the traditional VIKOR method and the improved VIKOR method are better than those of the TOPSIS method.

Further analysis of the basic performance of the sample compared the excess earnings performance of enterprises with different rankings. Evaluation accuracy was obtained by comparing the top 5 evaluation results with the median value of the excess return rate of all sample firms, while the bottom 5 results were compared with the median value to obtain the evaluation error rate. The validity test for the evaluation results of the small- and medium-sized construction enterprises listed in 2016 is shown in Table 7. The traditional VIKOR method and the improved VIKOR method are highly effective.

To verify the comparison of the dynamic growth types of enterprises, financial data from 2017 are considered. The growth types of enterprises in 2013–2016 and the growth rate of net profit in 2017 are shown in Table 8. The table shows that in 2017, the ranking of the average net profit growth rate of stable growth enterprises improved as follows: VIKOR method (43.69%) > TOPSIS method (28.05%) > traditional VIKOR method (−177.8%). The ranking of the average net profit growth rate of enterprises with negative growth (including stable negative growth and fluctuating negative growth) is as follows: VIKOR method (−31.39%) < TOPSIS method (−28.26%) < traditional VIKOR method (40.02%). These rankings show that the improved VIKOR method best distinguishes the advantages and disadvantages of enterprise growth.

Therefore, the improved VIKOR evaluation method proposed in this paper makes full use of the information for both the positive and negative ideal solutions to make the evaluation results more practical and reasonable.

3.3. Sensitivity Analysis

3.3.1. Parameter Sensitivity Analysis. To study the influence of the decision mechanism coefficient v of the improved VIKOR method on the growth evaluation of small- and medium-sized construction enterprises, the influence of model parameters on the ranking results of schemes is determined through sensitivity analysis [32]. Taking 2016 as an example, this section conducted a sensitivity analysis by selecting different v values. The value of parameter v is evaluated from the interval [0, 1] with a step length of 0.1 to analyse the growth sequencing of construction enterprises. As shown in Figure 1, the top company ranked first in all 11 experiments (100%), 8 companies (100%) ranked in the top 30% (top 8) in the 11 experiments, and 6 companies (75%) ranked in the bottom 30% (the last 8) in the 11 experiments.

The sensitivity analysis of the traditional VIKOR method to the coefficient V of the decision mechanism is shown in Figure 2. In the 11 experiments, the first-ranked enterprise ranked first in all experiments (100%), 5 enterprises (62.5%) ranked in the top 30% (top 8) in all experiments, and 4 (50%) enterprises ranked in the bottom 30% (bottom 8) in all experiments.

3.3.2. Sensitivity Analysis of Weights. When the original evaluation information objectively empowers each evaluation criterion and the weight changes, do the findings for an enterprise's growth change? What changes take place? Through sensitivity analysis, it can be determined that potential changes in the weight of the evaluation criteria lead to a change in the evaluation results, which is the key to effectively using the model and implementing quantitative decisions [32]. Taking 2016 as an example, this paper adopts a perturbation method to analyse the sensitivity of the weight of the evaluation criteria; that is, when the weight of the model evaluation criteria is slightly disturbed, the growth evaluation results for the sample enterprises may change accordingly. The initial weight of the evaluation criteria C_j is w_j; thereafter, a disturbance is denoted as $w_j' = \eta w_j\,(0 \le w_j' \le 1)$, and the variation interval of the parameter η is $0 \le \eta \le (1/w_j)$. According to the normalization of the weight, the other weight changes in accordance with the change in w_j, denoted as $w_k' = \varnothing w_k$, $k \ne j$, $k = 1, 2, \ldots, n$, and satisfies

Table 7: Effectiveness test of 2016 evaluation results.

	Top 5	Bottom 5	Evaluation accuracy (%)	Evaluation error rate (%)
TOPSIS method	2 (5)	3 (5)	40	60
Traditional VIKOR method	3 (5)	2 (5)	60	40
Improved VIKOR method	3 (5)	2 (5)	60	40

Table 8: Dynamic classification of enterprise growth in 2013–2016.

Methods	Growth type and average growth rate of net profit							
	Steady growth	Net profit growth rate	Fluctuating growth	Net profit growth rate	Steady negative growth	Net profit growth rate	Volatile negative growth	Net profit growth rate
TOPSIS method	7	28.05	1	−46.31	12	−79.48	6	74.18
Traditional VIKOR method	6	−177.8	1	−43.32	15	42.92	4	29.14
Improved VIKOR method	6	43.69	0		16	−46.52	4	29.14

$$w_k' + \sum_{k\neq i,k=1}^{n} w_k' = 1, \tag{17}$$

that is,

$$\eta w_j + \phi \sum_{k\neq i,k=1}^{n} w_k = 1. \tag{18}$$

The solution is $\phi = ((1-\eta w_j)/(1-w_j))$. For the weight of each evaluation criterion, when different parameters are adopted for perturbation, the corresponding evaluation results of the growth of construction enterprises are obtained by using the improved VIKOR method. The weights of the 25 evaluation indicators in this paper were changed, 1/2 and 1/3 were taken in turn, and 50 experiments were performed to obtain the analysis results, as shown in Figure 3. The top company ranked first in 48 of the 50 experiments (96%); the top 30% (top 8) companies ranked in the top 30% in 7 (87.5%) out of the 50 experiments, and 6 (75%) enterprises ranked in the bottom 30% (the bottom 8) in 50 experiments.

The sensitivity analysis of the traditional VIKOR method to weights is shown in Figure 4. The first-ranked enterprise ranked first (96%) in 48 of the 50 experiments. Five (62.5%) enterprises ranked in the top 30% (top 8) of all 50 experiments, and 6 (75%) enterprises ranked in the bottom 30% (bottom 8) of all 50 experiments.

The sensitivity analysis of the TOPSIS method to weights is shown in Figure 5. The first-ranked enterprise ranked first (100%) in the 50 experiments, 5 enterprises (62.5%) ranked in the top 30% (top 8) in all 50 experiments, and 6 (75%) enterprises ranked in the bottom 30% (bottom 8) in all 50 experiments.

3.3.3. Comparative Analysis. The sensitivity analysis of the three methods is shown in Table 9. In terms of parameter sensitivity, the average ranking consistency rate of the improved VIKOR method is 91.67%, which is higher than 70.83% of the traditional VIKOR method; in terms of weight sensitivity, the average value of the improved VIKOR method is 86.17%, which is higher than the 79.17% of the TOPSIS method and the 77.83% of the traditional VIKOR method. Therefore, the improved VIKOR method is more stable than the traditional VIKOR method and the TOPSIS method.

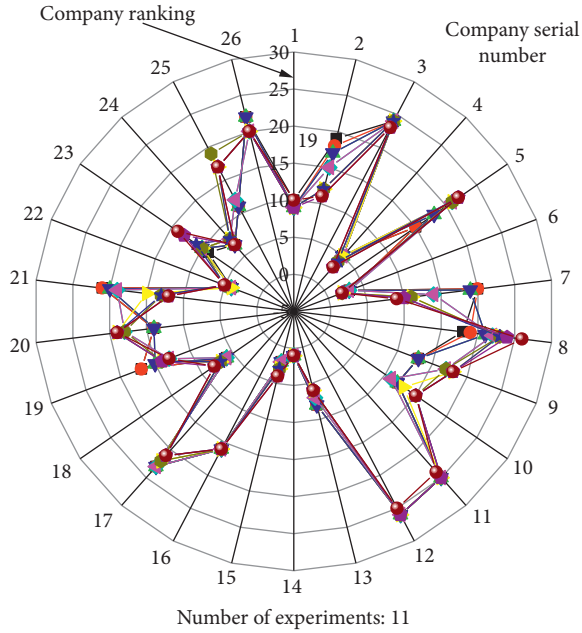


Figure 1: Improved VIKOR method parameter V sensitivity analysis results.

4. Discussion and Conclusions

(1) It can be seen from the index weight of each level (Table 3) that operation ability has the largest influence on enterprise growth. This finding is different from the finding in the literature [2] that the growth of listed companies in China's environmental protection industry is affected by profitability. The

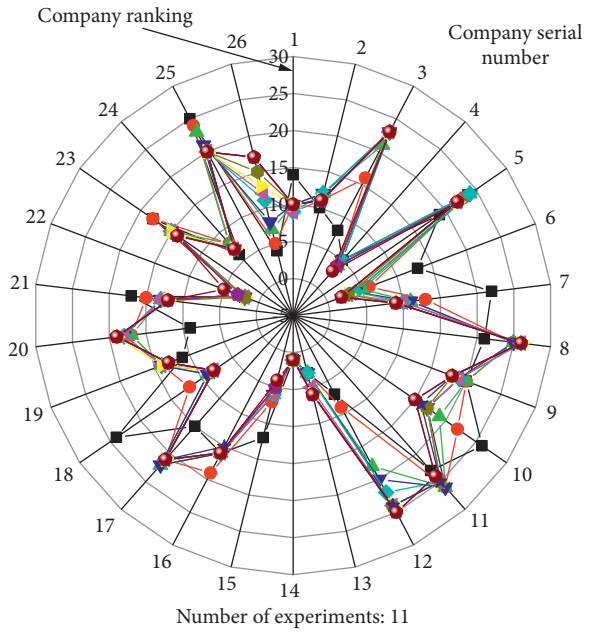


Figure 2: Traditional VIKOR parameter V sensitivity analysis results.

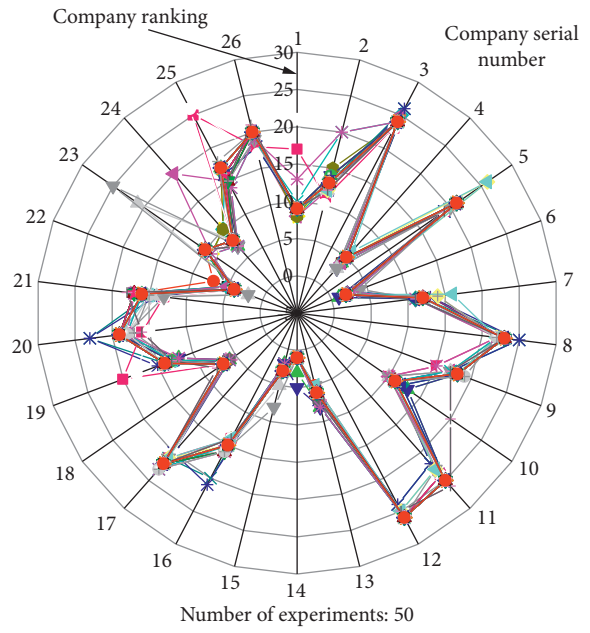


Figure 3: Weight sensitivity analysis results of the improved VIKOR method.

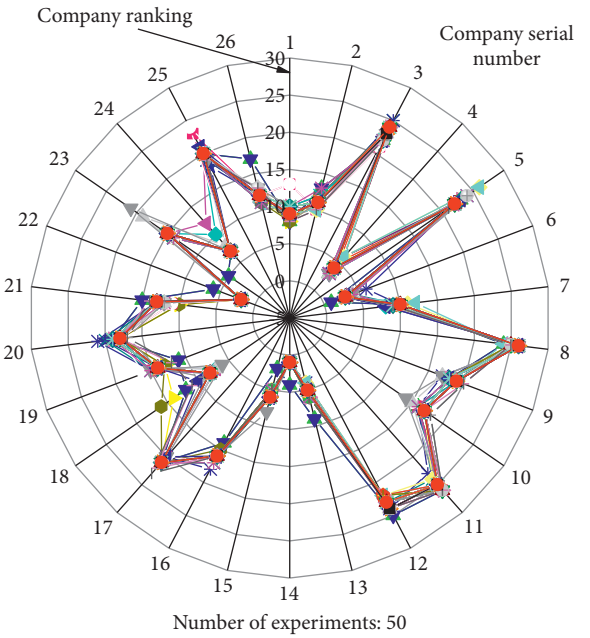


Figure 4: Weight sensitivity analysis results of the traditional VIKOR method.

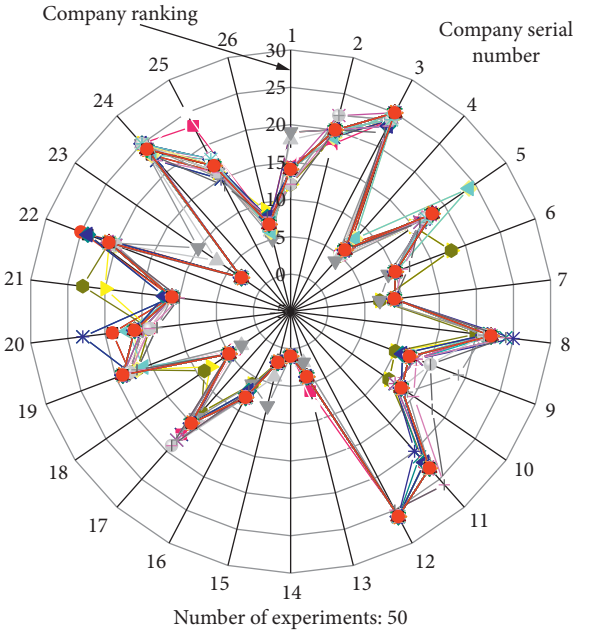


Figure 5: Weight sensitivity analysis results of the TOPSIS method.

difference in these findings is mainly due to the many uncertain factors, such as fierce bidding, engineering change, cost increases, inflation, and government policies that hinder the realization of earnings for construction enterprises. Therefore, construction enterprises should strengthen their operational capacity management and improve their operational management levels. They should focus on strengthening the management of operating capacity, intangible resources, tangible resources, human resources, and other key factors to improve growth.

(2) Regarding the comparison and analysis of the growth ranking of small- and medium-sized construction enterprises, Table 4 shows that ZTC, Golden Mantis, Oriental Garden, HONGTAO, Sanwei, and RECLAM are all ranked in the top 10 for growth between 2013 and 2016, while GHE, ZSSF, YT-ECO, and BXLQ are all ranked in the bottom 10. This

Table 9: Results of the sensitivity comparison.

Categories	Comparison items	TOPSIS method	Traditional VIKOR method	Improved VIKOR method
		Ranking consistency rate (%)		
Parameter sensitivity	Top 1		100	100
	Top 30%		62.5	100
	Bottom 30%		50	75
	Mean value		70.83	91.67
Weight sensitivity	Top 1	100	96	96
	Top 30%	62.5	62.5	87.5
	Bottom 30%	75	75	75
	Mean value	79.17	77.83	86.17

finding indicates that there is a serious imbalance in the growth performance of small- and medium-sized construction enterprises. Well-developed enterprises maintain good performance, while enterprises with poor development continually face difficulties. Therefore, enterprise managers and regulatory departments should conduct targeted management measures according to the ranking of the enterprise.

(3) Regarding the analysis of the dynamic growth of small- and medium-sized construction enterprises, Table 4 shows that there are 6 companies with stable growth, 16 with stable negative growth, and 4 with fluctuating negative growth. Only 23% (6) of the enterprises present positive growth, while 77% (20) present negative growth, indicating that the selected small construction enterprises have poor overall performance in terms of sustainable growth. This poor performance can be explained by several problems that are commonly faced by small- and medium-sized construction enterprises compared with large construction enterprises. As small- and medium-sized enterprises have weak market competitiveness, it is difficult for them to win bids for large projects, so they can rely only on large enterprises or engage in subcontracting projects. Small- and medium-sized construction enterprises have financing difficulties, declining economic benefits, and a low ability to withstand risks. They also face a serious shortage of high-level technical development and management talent, and family management is prominent. These enterprises are also limited by institutional issues, such as the division of the construction market and the qualification access system. The earnings of small- and medium-sized construction enterprises are affected by many uncertain factors. To face these challenges, such enterprises should seize opportunities, improve their independent innovation capabilities, innovate in project management, establish and improve modern enterprise systems based on a corporate governance structure, accelerate transformation and upgrading, transform their development mode, and achieve sustainable development.

(4) Regarding the effectiveness analysis of the improved VIKOR method, Tables 5 and 6 indicate that the correlation coefficients of the evaluation results of the TOPSIS method, traditional VIKOR method, and improved VIKOR method are high. The improved VIKOR method is shown to be effective and superior to the traditional VIKOR method in terms of the Q value. From Table 6, it can be seen that the evaluation results of the traditional VIKOR method and the improved VIKOR method are better than those of the TOPSIS method and are positively but nonsignificantly related to the annual excess return rate of stocks. This finding shows that stocks with higher growth rankings are more favoured by investors and have higher returns than those with lower growth rankings. However, because China's stock market is still an emerging market, speculation is strong, and value investment is difficult to promote; thus, this correlation is not significant. By further analysing the basic performance of the sample and comparing the different enterprises by their performance in terms of excess returns, it is concluded that the evaluation model has high effectiveness. Table 8 shows that the improved VIKOR method can distinguish the advantages and disadvantages of enterprise growth better than the TOPSIS method and the traditional VIKOR method, so its results are more practical. This finding shows that the evaluation model has good effectiveness.

(5) Regarding the sensitivity analysis of the improved VIKOR method, the ranking consistency rate reaches 75%, which shows that the ranking of small- and medium-sized construction enterprises with high growth and low rankings is not sensitive to the change in the V parameter. The ranking consistency rate also reaches 75% according to the weight sensitivity analysis of the evaluation model, which shows that the VIKOR method is not sensitive to weight disturbances.

In summary, based on the shortcomings of the traditional VIKOR method, this paper proposes an improved entropy-VIKOR algorithm that simultaneously considers information on positive and negative ideal solutions. It calculates the group utility value and the individual regret value by taking the positive and negative ideal solutions as references, respectively, and makes full use of the information on positive and negative ideal solutions, so that the

decision-making results are in line with reality. Based on this algorithm, the growth evaluation model of small- and medium-sized construction enterprises, including growth resources, growth ability, and growth innovation, is established. Through sensitivity and effectiveness analysis, it is concluded that the evaluation model proposed in this paper has better stability and effectiveness than the TOPSIS method and the traditional VIKOR method. The evaluation results are analysed, and several suggestions are put forward. The evaluation results analysed in this paper can provide a reference for managers, investors, and regulatory departments.

Data Availability

Data can be made available upon request.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this article.

Acknowledgments

This project was supported by the Project of Ecological Civilization Research Centre of Fujian Social Sciences Research Base in 2018 under grant KXJD1805A, Research Project in FAFU of General Secretary Xi Jinping's Important Thought and Exploration Practice in Fujian under grant XSX201805, National Scientific Research Project Cultivation Progress of Social Sciences in FAFU in 2018 under grant xpy201808, and Project of Fujian Soft Science under grant 2019R0036.

References

[1] N. Lu, Y. Q. Li, and J. Liu, "Study on appraisal model for the growth of small and medium-sized construction enterprises," *Construction Economy*, vol. 36, no. 3, pp. 5–9, 2015.

[2] Y. F. Wang, "Principal component analysis of enterprise growth: a case study of listed companies in environmental protection industry," *China Management Informationization*, vol. 13, no. 19, pp. 38–41, 2010.

[3] J. Hu, W. Wang, D. Jiang, and J. S. Yang, "The growth evaluation model and its application of fitness club of Jiangsu province," *Sports & Science*, vol. 1, pp. 55–62, 2018.

[4] S. Gong and Q. H. Lv, "Construction of index system and prediction of growth of sporting goods industry in China: data analysis based on R language," *Journal of Shenyang Sport University*, vol. 37, no. 3, pp. 15–22, 2018.

[5] S. L. Xia and N. N. Wang, "Growth evaluation of Jiangsu venture listed companies based on factor analysis method," *Science & Technology Management Research*, vol. 24, pp. 70–74, 2014.

[6] F. H. Gong and J. Gao, "Research on growth of small and medium manufacturing enterprises," *Journal of Xian Technological University*, vol. 1, pp. 46–51, 2013.

[7] D. Y. Shen, "Study on the endogenous growth of SMEs based on technological innovation capability," *Modern Economic Science*, vol. 3, pp. 116–123, 2017.

[8] H. F. Wang and Y. Zhao, "A comparative study on the growth of hi-tech enterprises in various provinces and cities of China based on the factor analysis method," *Science & Technology Management Research*, vol. 4, pp. 59–64, 2013.

[9] T. Lv and L. Pan, "Research on evaluation of the growth of new energy listed companies in China," *Journal of Industrial Technological & Economics*, vol. 2, pp. 118–125, 2017.

[10] Y. J. Zhu, "Growth evaluation index of small and medium-sized enterprise and empirical study," *Science & Technology Management Research*, vol. 13, pp. 75–78, 2013.

[11] Z. W. Zhao, P. Wang, and M. Y. Liu, "The growth evaluation research on high-tech enterprises based on dynamic cap," *Science & Technology Management Research*, vol. 23, pp. 74–77, 2016.

[12] H. M. Cao and Y. Chang, "Analysis of the effect of financial credit evaluation on the impact of growth," *On Economic Problems*, vol. 10, pp. 113–117, 2015.

[13] L. X. Zhang, P. Wang, and D. S. Liu, "Growth factors model building and empirical analysis of small and medium sized technology enterprises," *Science & Technology Management Research*, vol. 1, pp. 95–102, 2016.

[14] Z. Zeng, "A novel model for enterprise technological innovation capability evaluation with 2-tuple linguistic information," *Journal of Intelligent & Fuzzy Systems*, vol. 31, no. 1, pp. 541–546, 2016.

[15] Z. Liu, S. Hu, R. Mu, and L. Chen, "Research on the evaluation of growth quality of listed LED companies," *Light & Engineering*, vol. 3, pp. 232–238, 2017.

[16] T. Lahdelma and S. Laakso, "Network analysis as a method of evaluating enterprise networks in regional development projects," *Evaluation*, vol. 22, no. 4, pp. 435–450, 2016.

[17] F. Lefley and J. Sarkis, "Applying the FAP model to the evaluation of strategic information technology projects," *International Journal of Enterprise Information Systems*, vol. 1, no. 2, pp. 69–90, 2017.

[18] Y. J. Zhu, "Research on optimal selection of evaluation methods for innovative enterprises," *Journal of Anhui University (Philosophy and Social Sciences)*, vol. 6, pp. 143–147, 2009.

[19] A. M. Rugman and A. Verbeke, "Edith Penrose's contribution to the resource-based view of strategic management," *Strategic Management Journal*, vol. 23, no. 8, pp. 769–780, 2002.

[20] A. Mardani, E. Zavadskas, K. Govindan, A. A. Senin, and A. Jusoh, "VIKOR technique: a systematic review of the state of the art literature on methodologies and applications," *Sustainability*, vol. 8, no. 1, p. 37, 2016.

[21] A. Sanayei, S. Farid Mousavi, and A. Yazdankhah, "Group decision making process for supplier selection with VIKOR under fuzzy environment," *Expert Systems with Applications*, vol. 37, no. 1, pp. 24–30, 2010.

[22] K. Salehi, "A hybrid fuzzy MCDM approach for project selection problem," *Decision Science Letters*, vol. 1, pp. 109–116, 2015.

[23] K. Chatterjee and S. Kar, "Unified Granular-number-based AHP-VIKOR multi-criteria decision framework," *Granular Computing*, vol. 2, no. 3, pp. 1–23, 2017.

[24] R. Hall, "The strategic analysis of intangible resources," *Strategic Management Journal*, vol. 13, no. 2, pp. 135–144, 2006.

[25] D. Lavie, "Capability reconfiguration: an analysis of incumbent responses to technological change," *Academy of Management Review*, vol. 31, no. 1, pp. 153–174, 2006.

[26] H. Sun, X. P. Sun, and Z. Q. Fan, "An evaluation and empirical study on competitiveness of civil engineering construction enterprises in shanghai stock market," *Science & Technology Progress & Policy*, vol. 13, pp. 102–106, 2011.

[27] Y. J. Li, H. Y. Liu, and S. Y. Hao, "Competitiveness evaluation of construction industry of every province in China based on factor analysis and cluster analysis," *Journal of Huazhong University of Science and Technology*, vol. 2, pp. 78–83, 2014.

[28] G. W. Liu, F. Deng, and M. Wang, "Industrial competitiveness assessment for Chinese construction industry," *China Civil Engineering Journal*, vol. 7, pp. 157–164, 2011.

[29] R. X. Fan and M. J. Wang, "Analysis and evaluation of the financial quality of the listed building enterprise model," *Journal of Railway Science and Engineering*, vol. 3, pp. 695–701, 2015.

[30] A. Awasthi and G. Kannan, "Green supplier development program selection using NGT and VIKOR under fuzzy environment," *Computers & Industrial Engineering*, vol. 91, pp. 100–108, 2016.

[31] T. Z. Liu and Y. Yu, "Research on comprehensive evaluation of academic journal based on combined weight and VIKOR," *Mathematics in Practice and Theory*, vol. 49, no. 1, pp. 313–322, 2019.

[32] R. Simanaviciene and L. Ustinovichius, "Sensitivity analysis for multiple criteria decision making methods: TOPSIS and SAW," *Procedia—Social and Behavioral Sciences*, vol. 2, no. 6, pp. 7743-7744, 2010.

Hindawi
Discrete Dynamics in Nature and Society
Volume 2020, Article ID 1632986, 6 pages
https://doi.org/10.1155/2020/1632986

Research Article

Improving Processing Time for the Location Algorithm of Robots

Jing Chen and Liwen Chen

School of Information Science and Engineering, Fujian University of Technology, No. 33 Xueyuan Road University Town, Fuzhou, Fujian, China

Correspondence should be addressed to Liwen Chen; chenlw2002@163.com

Received 10 June 2020; Accepted 17 July 2020; Published 2 September 2020

Guest Editor: Chi-Hua Chen

The paper proposes an algorithm based on the Multi-State Constraint Kalman Filter (MSCKF) algorithm to construct the map for robots special in the poor GPS signal environment. We can calculate the position of the robots with the data collected by inertial measurement unit and the features extracted by the camera with MSCKF algorithm in a tight couple way. The paper focuses on the way of optimizing the position because we adopt it to compute Kalman gain for updating the state of robots. In order to reduce the processing time, we design a novel fast Gauss–Newton MSCKF algorithm to complete the nonlinear optimization. Compared with the performance of conventional MSCKF algorithm, the novel fast-location algorithm can reduce the processing time with the kitti datasets.

1. Introduction

The odometers of robots can divide into two parts: motion propagation and state updating end [1, 2]. The motion propagation is computing the motion matrix with the data of the sensors. Then, the result updates through the updating end. Many robots adopt the cameras to construct visual odometer because the cameras are conveniently mounted. [3] The visual odometers extract the matching features from the video frames to propagate the motion matrix. However, there are many drawbacks at the visual odometer such as the long processing time and the error because of mismatching features. In order to reduce the processing time, some researchers applied two threads to compute the motion equation and update the result, respectively [4, 5]. However, it will increase the hardware cost of the location systems by this way. In order to improve the processing time without additional hardware, Chansoo used the corner features to replace the ordinary features [6]. However, the information of the features sometimes will lose to make it hard to compute the location of robots. Moreover, it is difficult to extract the corner features in the texture flaw environment. In this way, the error of the system becomes big. In order to reduce the processing time and improve accuracy of visual odometer at the same time, more and more researchers began to adopt different sensors to compensate for the drawbacks of the visual odometers [7–12]. GPS can get the accurate location of the robot in the outdoor environment. Inertial measurement unit (IMU) can compensate for the error in the poor GPS signal environment [13, 14]. For the application of the robot in the poor GPS signal environment, many researchers adopted IMU to construct the visual inertial odometer (VIO) [15–19]. In order to compute the motion matrix and construct the map of robots by the VIO, we adopt the MSCKF algorithm. The motion propagation of MSCKF algorithm use the data of IMU to propagate the motion matrix. The update-end of MSCKF algorithm used the observed constraint features to update the motion matrix under the extended Kalman filter (EKF) framework [16, 20, 21]. However, the performance of conventional MSCKF about the processing time and the accuracy is bad. Ramezani et al. expanded the range of the observed features to improve the accuracy of the MSCKF algorithm [22]. Li et al. constructed the model of the error for the gravity accelerator and the bias error of the IMU to improve the accuracy of the MSCKF algorithm [23, 24]. Some researchers add the close-loop detection to update the initial point of the robot [24, 25]. Those methods improved the accuracy of the

system but increased the processing time [25]. In order to improve the processing time, Eckenhoff et al. proposed the online verification to reduce the update time and improve the accuracy of the MSCKF algorithm although it will increase the hardware cost [26]. Sejong Heo and others applied the slide windows to replace the ordinary method of extracting the features of MSCKF algorithm to reduce the computation amount and improve the processing time [17, 27, 28]. However, the processing time did not significantly reduce. Some researchers focused on the improvement of the back end of MSCKF algorithm to improve the processing time because the update end of the MSCKF algorithm based on the EKF filter will cost much process time. Some researchers applied the Cubature Kalman Filters (CKF) and Maximum Likelihood Estimate (MLE) method to replace the EKF although the performance of the processing time still is not ideal [29, 30]. Alibay et al. applied the stage Random Sample Consensus (RANSAC) method to propagate the motion matrix to meet the requirement of real time [31]. Nikolic applied Field Programmable Gate Array (FPGA) to complete the preprocessing the data of IMU to reduce the processing time [32]. However, those methods will increase the hardware cost.

In order to reduce the processing time, we design the novel fast Gauss–Newton MSCKF algorithm without additional hardware cost. The key point of our work is calculating the position and the attitude optimized by the novel nonlinear method. Then, we use the result to compute the Kalman gain in EKF framework.

2. Related Work

MSCKF algorithm is based on the EKF architecture. The algorithm is composed of three parts: propagation, augmentation, and update [16]. We use the state of IMU to propagate the state of robots at the state propagation. The propagation equation is as follows [16]:

$$\dot{\widetilde{X}}_{\text{imu}} = F\widetilde{X}_{\text{imu}} + \text{Gn}_{\text{imu}}, \tag{1}$$

where F means the coefficient matrix of the IMU state matrix that calculated by the accelerator matrix of IMU and the gyroscope data matrix and G means the coefficient matrix of the noise matrix. Then, we compute the Jacobin matrix to complete state augment. We define the matrix of the state of the No. j frame as the symbol $P_{j|j}$. The matrix is constructed of the covariance matrix of the state of the IMU ($P_{\text{II}_{j|j}}$), the covariance matrix between IMU and the camera ($P_{\text{IC}_{j|j}}$), and the covariance matrix of the cameras ($P_{\text{CC}_{j|j}}$):

$$P_{j+1|j} = \begin{bmatrix} P_{\text{II}_{j|j}} & P_{\text{IC}_{j|j}}J^T \\ \text{JP}_{\text{IC}_{j|j}}^T & \text{JP}_{\text{CC}_{j|j}}J^T \end{bmatrix}. \tag{2}$$

Considering the association between the features, MSCKF algorithm updates the propagation by the collection of the observed feature. At the same time, the MSCKF algorithm eliminates the features that are unobserved and worthless. Then, we can compute the reprojection errors of the observed features with nonlinear optimization to update the state matrix. Now, the observation equation denotes as follows:

$$r = H\widetilde{X} + \text{noise}, \tag{3}$$

where r means the residual error of the features, H denotes the Jacobin matrix, and $\widetilde{X}$ denotes the state propagated through the above motion equation. Now, we need to compute the matrix H to get the Kalman gain. The residual error can compute by the difference between the real position of the jth feature and the projected value of the position. The process denotes as follows:

$$\text{r}_i^{(j)} = z_i^j - \widehat{z}_i^j. \tag{4}$$

In this way, $\widehat{z}_i^j$ denotes the estimated position of the No. j feature observed by the No. i camera. z_i^j means the position of the No. j feature. In order to reduce the amount of computation, it does QR decomposition of $H_x^{(j)}$ to accelerate the solution of equations. Then, we get the upper triangular matrix T_H that is nonsingular matrix:

$$H_o = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} T_H \\ 0 \end{bmatrix}. \tag{5}$$

Now, we can use T_H to get the Kalman gain:

$$K = \text{PT}_H^T \left(T_H \text{PT}_H^T + R_{\text{n}}\right)^{-1}, \tag{6}$$

where P means the state matrix of the camera and R_n means the noise matrix. The state covariance of No. $j+1$ frame can be updated as follows:

$$P_{j+1|j+1} = \left(I_{6N+15} - \text{KT}_H\right) P_{j+1|j} \left(I_{6N+15} - \text{KT}_H\right)^T + \text{KR}_n K^T. \tag{7}$$

3. Our Method

The value of $\widehat{z}_i^j$ determines the result of the Kalman gain K in equation 6. Calculating the $\widehat{z}_i^j$ costs the most processing time, the conventional method adopts the Gauss–Newton optimization method to compute the value of ${}^G\widehat{p}_{f_j}$ that used to calculate $\widehat{z}_i^j$. In order to reduce the processing time, we employ the fast Gauss–Newton method to make the nonlinear optimization. The whole process of optimization is divided into two stages.

At the first stage, the difference is so big that we use the gradient descent method to optimize. We make the Taylor expansion of the equation as follows:

$$f(x + \Delta x) \approx f(x) + J(x)\Delta x. \tag{8}$$

We can get the gradient descent of Δx is $J(x)$. However, the gradient descent method may cause the diverged result because iterative step size is too big.

At the second stage, we use the Gauss–Newton method to get Δx when the optimized value is very close to the prior value to keep the convergence of algorithm. The objection of Gauss–Newton algorithm is to minimize $\|f(x + \Delta x)\|^2$.

So, we get the solution of the derivation. We get the Gauss Newton equation:

$$J(x)^T J(x)\Delta x = -J(x)^T f(x). \tag{9}$$

Then, we can compute the value of Δx as follows:

$$\Delta x = -\left(J(x)^T J(x)\right)^{-1} J(x)^T f(x). \tag{10}$$

In order to compute the real value of motion matrix, we bring the prior value of the coordinate into the above equations. If we describe the prior-estimated value of position of the No. j frame $({}^{C_i}\widehat{p}_{f_j})$ as follows:

$${}^{C_i}\widehat{p}_{f_j} = C\left({}^{C_i}_{C_n}\widehat{\overline{q}}\right)\begin{bmatrix} {}^{C_n}\widehat{X}_j \\ {}^{C_n}\widehat{Y}_j \\ {}^{C_n}\widehat{Z}_j \end{bmatrix} + {}^{C_i}p_{c_n}, \tag{11}$$

where ${}^{C_i}\widehat{p}_{f_j}$ means the prior-estimated value of the position of the jth frame and $C({}^{C_i}_{C_n}\widehat{\overline{q}})$ means the matrix of rotation quaternion We can change the above equation by normalization as follows:

$${}^{C_i}\widehat{p}_{f_j} = {}^{C_n}\widehat{Z}_j \left(C\left({}^{C_i}_{C_n}\widehat{\overline{q}}\right)\begin{bmatrix} \dfrac{{}^{C_n}\widehat{X}_j}{{}^{C_n}\widehat{Z}_j} \\ \dfrac{{}^{C_n}\widehat{Y}_j}{{}^{C_n}\widehat{Z}_j} \\ 1 \end{bmatrix} + \frac{1}{{}^{C_n}\widehat{Z}_j}\,{}^{C_i}p_{C_n} \right). \tag{12}$$

In order to make clearer of the expression, we describe as follows:

$$\begin{aligned} \widehat{\alpha}_j &= \frac{{}^{C_n}\widehat{X}_j}{{}^{C_n}\widehat{Z}_j}, \\ \beta_j &= \frac{{}^{C_n}\widehat{Y}_j}{{}^{C_n}\widehat{Z}_j}, \\ \widehat{\rho}_j &= \frac{1}{{}^{C_n}\widehat{Z}_j}. \end{aligned} \tag{13}$$

Now, we can update ${}^{C_i}\widehat{p}_{f_j}$ as follows:

$${}^{C_i}\widehat{p}_{f_j} = {}^{C_n}Z_j \begin{bmatrix} f_{i1}\left(\widehat{\alpha}_j, \widehat{\beta}_j, \widehat{\rho}_j\right) \\ f_{i2}\left(\widehat{\alpha}_j, \widehat{\beta}_j, \widehat{\rho}_j\right) \\ f_{i3}\left(\widehat{\alpha}_j, \widehat{\beta}_j, \widehat{\rho}_j\right) \end{bmatrix}. \tag{14}$$

Now, we update the error between the estimated value and the prior value as follows:

$$\begin{aligned} e(\widehat{\sigma}) &= \widehat{z}_i^{(j)} - \frac{1}{f_{i3}(\widehat{\sigma})}\begin{bmatrix} f_{i1}(\widehat{\sigma}) \\ f_{i2}(\widehat{\sigma}) \end{bmatrix}, \\ \widehat{\sigma} &= \begin{bmatrix} \widehat{\alpha}_i \\ \widehat{\beta}_i \\ \widehat{\rho}_j \end{bmatrix}. \end{aligned} \tag{15}$$

The prior value of $\widehat{z}_i^{(j)}$ is normalization value in this. We can use J_k to stand for the effect of the iteration. We describe J_k as follows:

$$J_k = 0.5 * e(\widehat{\sigma})^T R^{-1} e(\widehat{\sigma}), \quad R = \left\{R^1, \ldots, R^n\right\}. \tag{16}$$

In the above equation, R denotes the weight of noise which includes the bias of gyroscope and the bias of accelerator velocity. In order to reflect the effect of the derivation, we describe the difference between the prior value and the optimized value as η:

$$\eta = \frac{\left|J_{k+1} - J_k\right|}{J_k}. \tag{17}$$

At the first stage, when η is bigger than the threshold value μ, we choose gradient descent as the value of Δx. By the way, the value of μ is the empirical value as 0.5. Now, we describe the step of the derivate δy as follows:

$$\delta y = -J(x)^T e(\widehat{y}), \quad \eta > \mu, \tag{18}$$

where $J(x)$ means the Jacobian matrix of the error $e(\widehat{y})$. At the second stage, when the optimized value is close to the real ones that means the value of η is less than μ. In order to guaranteed convergence of matrix, we choose Gauss–Newton algorithm to complete nonlinear optimization. The step of Δx denotes as follows:

$$\delta y = \left(J^T R^{-1} J\right)^{-1}\left(-J^T R^{-1}\right) e(\widehat{y}), \quad 0.01 < \eta <= \mu. \tag{19}$$

If η is smaller than 0.01, the algorithm converges, and we can stop iteration.

4. Experiment

The experiment of the paper based on kitti datasets. The preprocess of the datasets is extracting the features and checking the external parameters of the cameras [33]. In the processing of the computation, we choose the value of the bias of acceleration and the gyroscope as 10^{-6}. We also choose the same value as the noise of the accelerator and the gyroscope. The result are shown in Table 1.

There are 98 frames of No. 1 datasets at Table 1. Compare the processing time of the conventional MSCKF algorithm, and we found that the processing time decreased by 20.7% when we adopt the novel fast Gauss–Newton MSCKF algorithm. Then, we calculate the final errors by the difference between the calculated result by different algorithms and the real value. Comparing the final errors of the MSCKF algorithm from the IMU data alone, the final error of conventional MSCKF decreases by 56.5%. The final error of the novel fast Gauss–Newton MSCKF decreases by 55.1%. The

Table 1: The performance of the novel MSCKF algorithm with No.1 datasets.

	Process time (s)	Final error (m)
Conventional MSCKF with the data of IMU alone	19.02	2.32
Conventional MSCKF with the data of IMU and the camera	19.02	1.01
Fast Gauss–Newton MSCKF with the data of IMU and the camera	15.09	1.04

Table 2: The performance of improved MSCKF algorithm with No. 2 datasets.

	Process time (s)	Final error (m)
Conventional MSCKF with the data of IMU alone	58.80	2.29
Conventional MSCKF with the data of IMU and the camera	58.80	3.32
Fast Gauss–Newton MSCKF with the data of IMU and the camera	55.30	3.52

final error calculated by the novel MSCKF algorithm increases by 3.0% than the one calculated by the conventional MSCKF algorithm.

There are 239 frames of No. 2 datasets at Table 2. Comparing the processing time of conventional MSCKF algorithm, the processing time of the novel fast Gauss–Newton MSCKF decreased by 6.0%. Comparing the final error of conventional MSCKF with the data of IMU alone, the final error of conventional MSCKF increased by 55.0%. The final error of the novel Gauss–Newton MSCKF algorithm increased by 6.0% relative to the final error of the conventional MSCKF algorithm.

Then, we can compare the trajectory computed by the conventional MSCKF algorithm with the one computed by the novel MSCKF algorithm. In order to make clear the result, we also compare the trajectory computed by the conventional MSCKF using the data of IMU alone with the global truth trajectory.

There are the trajectories calculated by different algorithms. The blue trajectory calculated by the conventional MSCKF algorithm is shown in Figure 1. The blue one of Figure 2 means the trajectory calculated by the novel Gauss–Newton MSCKF algorithm. There are red trajectories calculated by the conventional MSCKF with the data IMU alone at both figures. There are the global truth trajectories in green color at both figures. We can get the conclusion that the blue trajectory is closer to the red trajectory than the green trajectory in Figure 1. It means the conventional MSCKF algorithm can fuse the data from IMU and the stereo camera to reduce the error. The blue trajectory is also closer to the green trajectory than the red trajectory in Figure 2. It means the novel MSCKF algorithm is also able to fuse the data of IMU and the features of the cameras to reduce the error calculated by the MSCKF with the data of the IMU alone. When we compared the blue trajectories at both figures, the blue trajectory of Figure 2 is closer to the red one than the blue trajectory of Figure 1 at the beginning. However, the error of the novel algorithm becomes bigger when the distance becomes longer at Figure 2. We also notice that there is a sudden change in the middle of the trajectory in Figure 2 because there is the switching between two optimization methods.

We also calculate the trajectories by the above algorithms with No. 2 datasets to test the performance of the novel MSCKF algorithm when the VIO runs longer. The red

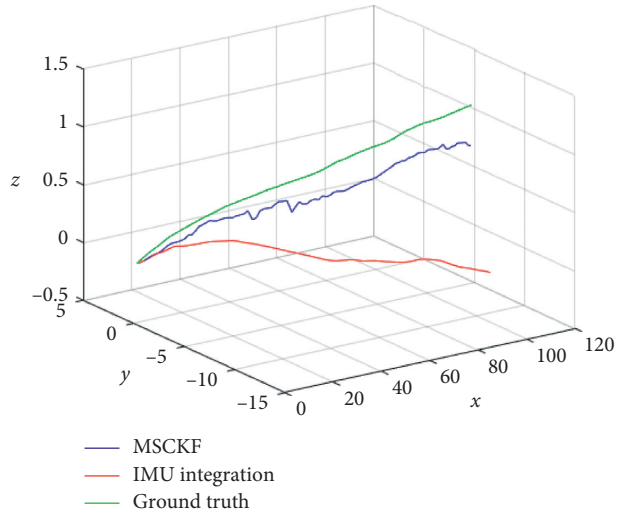


Figure 1: The trajectories by conventional MSCKF algorithm with No. 1 datasets.

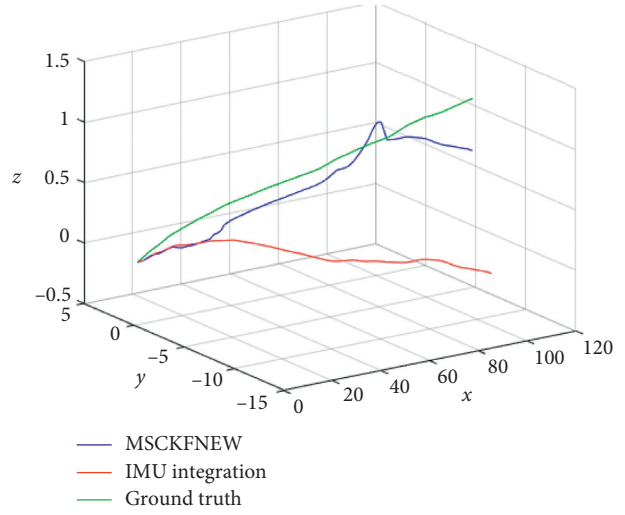


Figure 2: The trajectories by novel MSCKF algorithm with No.1 datasets.

trajectories means the ones calculated by the conventional MSCKF algorithm with the IMU data alone under No. 2 datasets in Figure 3 and 4. The green trajectories means the

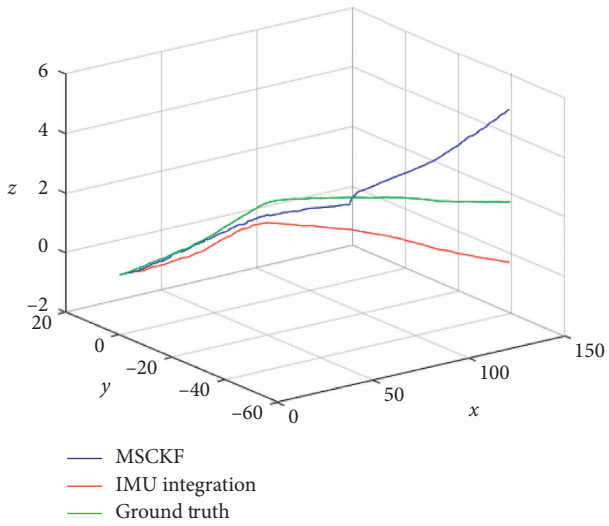


Figure 3: The trajectories of MSCKF with No. 2 datasets.

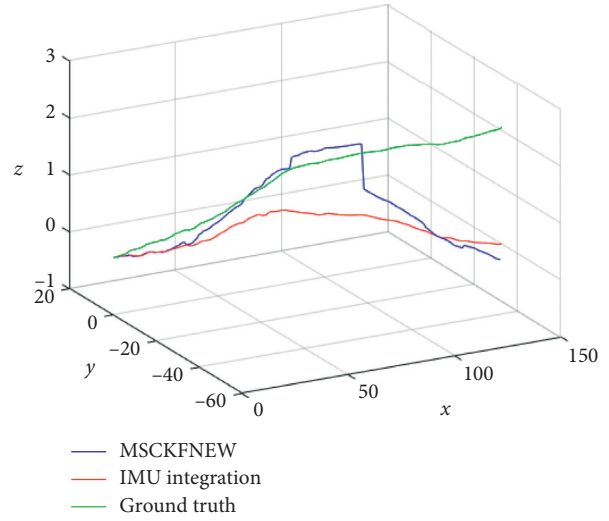


Figure 4: The trajectories of improved MSCKF with No. 2 datasets.

ground truth trajectory in Figures 3 and 4. The blue trajectory of Figure 3 means the trajectory calculated by the conventional MSCKF algorithm with No. 2 datasets. The blue one of Figure 4 means the trajectory calculated by the novel Gauss–Newton MSCKF algorithm with No. 2 datasets. We notice that the blue trajectory at Figure 3 is closer to the green one at first. However, the blue trajectory becomes deviated from the red trajectory when the distance becomes longer at Figure 3. It means the error of novel MSCKF algorithm becomes bigger when the VO runs longer. The blue trajectory becomes closer to the red trajectory at Figure 4 at first. However, when the distance x is between 50 m and 100 m, the blue trajectory is closer to the green one. It means the error computed by the novel MSCKF decrease when the computation continues. When the distance x is bigger than 100 m the error of blue trajectory became bigger. Nevertheless, there is more obvious hip change than the blue trajectory in Figure 2.

5. Conclusion

Comparing the performances of the novel MSCKF algorithm with the conventional MSCKF algorithm, the processing time of the novel MSCKF algorithm decreases by 20.7%~6.0%. The final error calculated by the novel MSCKF algorithm increases by 2.0%~6.0% than the one calculated by the conventional MSCKF algorithm. We can get the conclusion that the novel MSCKF can reduce the processing time. However, the final error computed by the fast Gauss–Newton MSCKF becomes bigger when the running time is longer. It means the novel MSCKF will increase the final error. We can employ it at the low precision and fast speed occasion. Then, there is obvious hip change at the trajectories calculated by the novel MSCKF algorithm because of the switching of the two optimization methods. We also find that the error calculated by the novel MSCKF becomes bigger when the VIO runs longer. In order to reduce the accumulated error, we can add the close-loop detection.

Data Availability

The data of the manuscript is from the kitti datasets which can be accessed through MATLAB. The data can be obtained from https://figshare.com/s/395bc65e815a4e4b3f2f.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Department of the Fujian Science and Technology (Grant 2018H0003), Fujian University of Technology (Grant GY-Z17011), Quanzhou Science and Technology Bureau (Grant 2017G012), Department of Education of Fujian Province (Grant JT180339), and Fujian Universtiy of Technology (JGBK201911).

References

[1] x. Gao, *14 Lessons of Visual Slam: From Theory to Practice*, Beijing, China, 2007.

[2] Y. Sun, "Overview of visual inertial SLAM," *The Application Research of Computers*, vol. 36, no. 12, pp. 3530–3533, 2019.

[3] D. Nistér, O. Naroditsky, and J. Bergen, "Visual odometry for ground vehicle applications," *Journal of Field Robotics*, vol. 23, no. 1, pp. 3–20, 2006.

[4] G. Klein and D. Murray, "Parallel tracking and mapping for small AR workspaces," in *Proceedings of the 2017th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 225–234, Nara, Japan, October 2007.

[5] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, "ORB-SLAM: a versatile and accurate monocular SLAM system," *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.

[6] P. Chansoo, "Illumination change compensation and extraction of corner feature orientation for upward-looking camera-based SLAM," in *Proceeedings of the 12th*

International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), pp. 224–227, Goyang, Korea, May 2015.

[7] S. Kohlbrecher and O. von Stryk, "A flexible and scalable SLAM system with full 3D motion estimation," in *Proceedings of the IEEE International Symposium on Safety, Security, and Rescue Robotics*, pp. 155–160, Kyoto, Japan, September 2011.

[8] X. Zuo, P. Geneva, Y. Yang, W. Ye, Y. Liu, and G. Huang, "Visual-inertial localization with prior LiDAR map constraints," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 3394–3401, 2019.

[9] S.-c. Chu and X. xue, "Optimizing ontology alignment in vector space," *Journal of Internet Technology*, vol. 21, no. 1, pp. 15–23, 2020.

[10] X. Xue and J. Chen, "Efficient user involvement in semi-automatic ontology matching," in *Proceedings of the IEEE Transactions on Emerging Topics in Computational Intelligence*, pp. 1–11, Piscataway, NJ, USA, December 2018.

[11] X. Xue and J. Chen, "Using compact evolutionary tabu search algorithm for matching sensor ontologies," *Swarm and Evolutionary Computation*, vol. 48, pp. 25–30, 2019.

[12] C.-H. Chen, F.-J. Hwang, and H.-Y. Kung, "Travel time prediction system based on data clustering for waste collection vehicles," *IEICE Transactions on Information and Systems*, vol. E102.DD, no. 7, pp. 1374–1383, 2019.

[13] K. Sun, K. Mohta, B. Pfrommer et al., "Robust stereo visual inertial odometry for fast autonomous flight," *IEEE Robotics and Automation Letters*, vol. 3, no. 2, pp. 965–972, 2018.

[14] A. Yilmaz and A. Gupta, "Indoor positioning using visual and inertial sensors," in *Proceeding of the 2016 IEEE Sensors*, pp. 1–3, Orlando, FL, USA, November 2016.

[15] N. N. Win, "A novel particle filter based SLAM algorithm for lunar navigation and exploration," in *Proceedings of the International Conference on Robotics and Automation Engineering (ICRAE)*, pp. 74–78, Montreal, Canada, May 2019.

[16] A. I. Mourikis and S. I. Roumeliotis, "A multi-state constraint kalman filter for vision-aided inertial navigation," in *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 3565–3572, Roma, May 2007.

[17] L. E. Clement, V. Peretroukhin, J. Lambert, and J. Kelly, "The battle for filter supremacy: a comparative study of the multi-state constraint kalman filter and the sliding window filter," in *Proceedings of the 12th Conference on Computer and Robot Vision*, pp. 23–30, Halifax, NS, USA, June 2015.

[18] N. Qi, "An improved MSCKF algorithm based onmulti-mode augmentation method for the camera state equation," *Chinese Journal of Scientific Instrument*, vol. 40, no. 05, pp. 89–98, 2019.

[19] Y. Sun, "Method of indoor robot positioning based on improved," *MSCKF Algorithm", Computer Systems Applications*, vol. 29, no. 2, pp. 238–243, 2020.

[20] G. Jianjun and G. Dongbing, "A direct visual-inertial sensor fusion approach in multi-state constraint Kalman filter," in *Proceedings of the 34th Chinese Control Conference (CCC)*, pp. 6105–6110, Hangzhou, China, July 2015.

[21] T. Nguyen, G. K. I. Mann, A. Vardy, and R. G. Gosine, "Likelihood-based iterated cubature multi-state-constraint Kalman filter for visual inertial navigation system," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4410–4415, Vancouver, BC, USA, November 2017.

[22] M. Ramezani, K. Khoshelham, and L. Kneip, "Omnidirectional visual-inertial odometry using multi-state constraint Kalman filter," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1317–1323, Vancouver, BC, USA, November 2017.

[23] J. Li, H. Bao and G. Zhang, Rapid and robust monocular visual-inertial initialization with gravity estimation via vertical edges," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 6230–6236, Macau, China, November 2019.

[24] P. Geneva, K. Eckenhoff, and G. Huang, "A linear-complexity EKF for visual-inertial navigation with loop closures," in *Proceedings of the International Conference on Robotics and Automation (ICRA)*, pp. 3535–3541, Montreal, QC, Canada, March 2019.

[25] M. Li and A. I. Mourikis, "Improving the accuracy of EKF-based visual-inertial odometry," in *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 828–835, Saint Paul, MN, Canada, May 2012.

[26] K. Eckenhoff, P. Geneva, J. Bloecker, and G. Huang, "Multi-camera visual-inertial navigation with online intrinsic and extrinsic calibration," in *Proceedings of the International Conference on Robotics and Automation (ICRA)*, pp. 3158–3164, Montreal, QC, Canada, March 2019.

[27] S. Heo, J. Cha, and C. G. Park, "EKF-based visual inertial navigation using sliding window nonlinear optimization," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 7, pp. 2470–2479, 2019.

[28] J. Hu and M. Chen, "A sliding-window visual-IMU odometer based on tri-focal tensor geometry," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3963–3968, Hong Kong. China, March 2014.

[29] K. Wen, W. Wu, X. Kong, and K. Liu, "A comparative study of the multi-state constraint and the multi-view geometry constraint kalman filter for robot ego-motion estimation," in *Proceedings of the eighth International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, Hangzhou, China, September 2016.

[30] J. Xu, H. Yu, and R. Teng, "Visual-inertial odometry using iterated cubature Kalman filter," in *Proceedings of the Chinese Control And Decision Conference (CCDC)*, pp. 3837–3841, Shenyang, China, June 2018.

[31] M. Alibay, S. Auberger, B. Stanciulescu, and P. Fuchs, "Hybrid visual and inertial RANSAC for real-time motion estimation," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pp. 179–183, Paris, France, July 2014.

[32] J. Nikolic, "A synchronized visual-inertial sensor system with FPGA pre-processing for accurate real-time SLAM," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pp. 431–437, Hong Kong, China, June 2014.

[33] L. E. Clement, V. Peretroukhin, J. Lambert, and J. Kelly, "The battle for filter supremacy: a comparative study of the multi-state constraint kalman filter and the sliding window filter," in *Proceedings of the 12th Conference on Computer and Robot Vision*, pp. 23–30, Halifax, Canada, June 2015.

Hindawi
Discrete Dynamics in Nature and Society
Volume 2020, Article ID 6787608, 9 pages
https://doi.org/10.1155/2020/6787608

Research Article

A Novel THz Differential Spectral Clustering Recognition Method Based on t-SNE

Tie-Jun Li,[1] Chih-Cheng Chen,[1,2] Jian-jun Liu,[3] Gui-fang Shao,[4] and Christopher Chun Ki Chan[5]

[1]*College of Information Engineering, Jimei University, Fujian Province, Xiamen 361021, China*
[2]*Department of Aeronautical Engineering, Chaoyang University of Technology, Taichung, Taiwan*
[3]*Shaoguan University, Guangzhou Province, Shaoguan 512005, China*
[4]*Department of Automation, Xiamen University, Fujian Province, Xiamen 361005, China*
[5]*Department of Information Management, Chaoyang University of Technology, Taichung, Taiwan*

Correspondence should be addressed to Chih-Cheng Chen; 3343033397@qq.com and Gui-fang Shao; gfshao@xmu.edu.cn

Received 6 March 2020; Accepted 6 July 2020; Published 1 September 2020

Guest Editor: Fuqiang Gu

We apply time-domain spectroscopy (THz) imaging technology to perform nondestructive detection on three industrial ceramic matrix composite (CMC) samples and one silicon slice with defects. In terms of spectrum recognition, a low-resolution THz spectrum image results in an ineffective recognition on sample defect features. Therefore, in this article, we propose a spectrum clustering recognition model based on t-distribution stochastic neighborhood embedding (t-SNE) to address this ineffective sample defect recognition. Firstly, we propose a model to recognize a reduced dimensional clustering of different spectrums drawn from the imaging spectrum data sets, in order to judge whether a sample includes a feature indicating a defect or not in a low-dimensional space. Second, we improve computation efficiency by mapping spectrum data samples from high-dimensional space to low-dimensional space by the use of a manifold learning algorithm (t-SNE). Finally, to achieve a visible observation of sample features in low-dimensional space, we use a conditional probability distribution to measure the distance invariant similarity. Comparative experiments indicate that our model can judge the existence of sample defect features or not through spectrum clustering, as a predetection process for image analysis.

1. Introduction

Nondestructive testing is one of the most significant applications of terahertz technology, and terahertz time-domain spectroscopy (THz—TDS) system is a commonly used technique [1, 2]. Information from each pixel of spectrum analysis can be obtained by terahertz determinant scanning and analysis of a large number of high-dimensional spectral image data to detect superficial damage and internal defects such as bubbles, cracks, and impurities of samples. However, when applying terahertz nondestructive testing, a terahertz wave is usually subject to a diffraction limit and a spatial optical resolution limit of the system, which makes it difficult to identify target microdefect structures with an optical resolution lower than terahertz. To solve this kind of problem, improving the precision of optical hardware and image processing [3, 4] is a common method but with limited effectiveness. Since the spectrum of the defect feature of a detected target sample is a "differential spectrum," (different than the normal structure), the cluster of the abnormal spectrum can be identified in a low-dimensional space through unsupervised clustering using the terahertz spectral data set from the perspective of spectral clustering recognition.

To observe the distribution of sample points from high-dimensional spectral data, it is necessary to reduce the dimensionality of it [5]. We can then perform terahertz spectral data redundancy and noise removal through the extraction of the main spectral features of each local point. The sample points cluster in two or three dimensions. At

present, the manifold learning method is usually used in dimension reduction for hyperspectral data clustering recognition, along with principal components analysis (PCA), multidimensional scaling analysis (MDS) [6], isometric feature mapping (ISOMAP) [7], locally linear embedding (LLE) [8], and spectrum embedding (SE) [9]. However, these data dimensionality reduction methods have problems such as unclear classification interface, poor visual classification effect, and slow convergence speed.

Maaten proposed a t-distribution SNE (t-SNE) algorithm [10, 11] based on the stochastic neighbor embedding (SNE) algorithm [12], which is a nonlinear dimension reduction method based on manifold learning. According to the principle that data points with similar distances in high-dimensional space are mapped to low-dimensional space with similar distances, a method of subspace analysis is adopted to measure the similarity of such spatial distances with conditional probability distribution. t-SNE changes the idea of similarity based on Gaussian distance or Euclidean distance measure in MDS and ISOMAP algorithms. It maps high-dimensional space sample points to low-dimensional space; meanwhile, the distribution probability between them remains unchanged as far as possible. Because the t-SNE algorithm has excellent visual classification effect, clear classification interface, and high algorithm efficiency, the method has been widely used in biomedical data analysis, fault diagnosis, spectrum analysis, artificial intelligence, and many other fields. For example, applied t-SNE dimensional reduction techniques are used to classify disease cells [13, 14], human genetic patterns [15], and RNA sequences [16]. By utilizing the t-SNE technique, other literatures [16–18] also applied it to classify multifaults in mechanical systems. In order to realize the classification and visualized diagnosis on the spectral information, the authors in [17–19] also made relevant research progress in their respective fields. In recent years, with the rapid development of artificial intelligence, the t-SNE technique was also applied to different AI-related applications [20–22].

This paper proposes a spectral identification model based on the t-SNE algorithm for the "differential spectrum" of sample defect features as a nondestructive method of testing. Based on the unsupervised clustering of "differential spectrum" from terahertz spectral data, we can achieve super-resolution identification of sample defect features at the pixel level, thus performing predetection analysis for further terahertz spectral imaging.

2. Terahertz Spectral Recognition Model Based on t-Distributed Stochastic Neighbor Embedding

The establishment of t-SNE terahertz spectral recognition model consists of the following four steps:

(1) Define the data set, calculate the confusion cost function, and initialize the optimization parameters of the model

(2) Set the low-dimensional data representation of optimized results

(3) Obtain target results from stochastic gradient descent optimization training

(4) Iterate through the pipeline until the number of iterations is reached

The algorithm flowchart is shown in Figure 1.

2.1. Basic Parameters Definition. To build the t-SNE model, we first define some basic parameters.

Let X be the spectral data set in a higher dimensional space, X_i represents the sample point and $X = \{x_1, x_2, \ldots, x\}$, and the dimension of the sample is D. Low-dimensional spatial data sets are represented by $Y = \{y_1, y_2, \ldots, y_n\}$, and the dimension of the sample is d with value of 2 or 3 to visualize the cluster analysis. The conditional profile distribution matrix P_i of a high-dimensional data set is defined as follows:

$$p_{j|i} = \frac{\exp\left(-\left\|x_i - x_j\right\|^2 / 2\sigma_i^2\right)}{\sum_{k \neq i} \exp\left(-\left\|x_i - x_k\right\|^2 / 2\sigma_i^2\right)}, \tag{1}$$

in which $P_{j|i}$ represents the probability that the ith sample is distributed around sample j, $P_{j|i} = 0$; σ denotes the variance of the Gaussian distribution centered on x_i and is determined according to the principle of maximum entropy. The entropy $H(P_i)$ in which P_i increases with the increase in σ_i, is defined as

$$H(P_i) = -\sum_j p_{j|i} \log_2 p_{j|i}. \tag{2}$$

To evaluate the number of effective nearest neighbors around a point, we introduce the concept of perplexity, which is a global parameter and defined as follows:

$$\text{Perp}\left(P_i\right) = 2^{H(P_i)}. \tag{3}$$

In order to make the adjustment of perplexity more robust, the perplexity is usually chosen between 5 and 50, and the binary search method is used to find the best σ.

The conditional probability distribution matrix Q_i in low-dimensional space is defined as follows:

$$q_{j|i} = \frac{\exp\left(-y_i - y_j^2\right)}{\sum_{k \neq i} \exp\left(-y_i - y_k^2\right)}. \tag{4}$$

2.2. Symmetric t-SNE. We let the probability distribution matrix in high-dimensional and low-dimensional space be symmetric and construct the joint probability distribution P and Q so that for any i and j, $p_{ij} = p_{ji}$ and $q_{ij} = q_{ji}$.

We redefine q_{ij} in low-dimensional space by t-distribution:

$$q_{ij} = \frac{\left(1 + y_i - y_j^2\right)^{-1}}{\sum_{k \neq i} \left(1 + y_k - y_l^2\right)^{-1}}. \tag{5}$$

Then, define q_{ij} in higher dimensions:

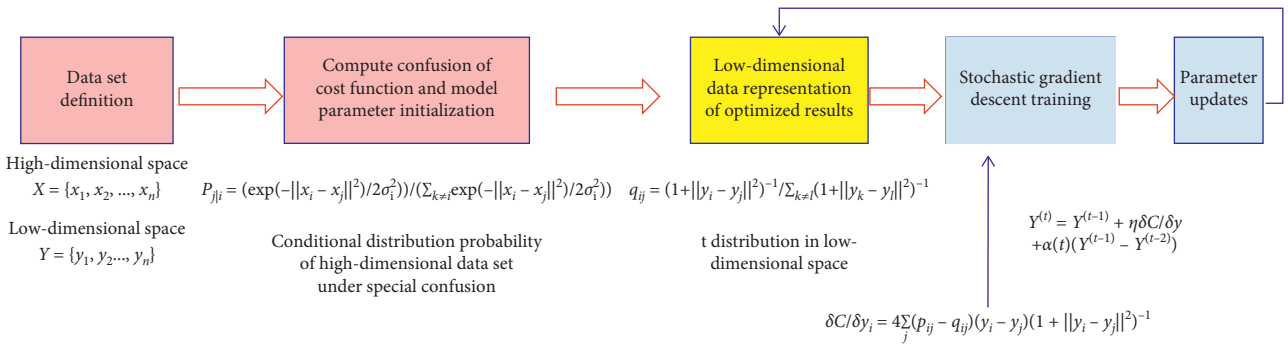


Figure 1: The flowchart of spectrum recognition based on the t-SNE algorithm.

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2n}, \tag{6}$$

in which n is the total number of sample points in the data set.

2.3. Cost Function and Training. We use Kullback–Leibler divergence (KLD) to measure the similarity of two spatial distributions in high and low dimensions, and the SEN algorithm aims to minimize the KL distance for all data points in the sample set.

We then use gradient descent to minimize cost function:

$$C = KL(PQ) = \sum p_{ij}\log_2\frac{p_{ij}}{q_{ij}}. \tag{7}$$

The gradient descent is also used for training the model and its formula is as follows:

$$\frac{\delta C}{\delta y_i} = 4\sum_j \left(p_{ij} - q_{ij}\right)\left(y_i - y_j\right)\left(1 + \left\|y_i - y_j\right\|^2\right)^{-1}. \tag{8}$$

In addition, to accelerate the optimization process and avoid falling just obtaining a local optima, a relatively large momentum should be used in the gradient; that is, in addition to the current gradient, the exponential decay term accumulated by the previous gradient should also be introduced in the parameter update. The formula is as follows:

$$Y^{(t)} = Y^{(t-1)} + \eta\frac{\delta C}{\delta y} + \alpha(t)\left(Y^{(t-1)} - Y^{(t-2)}\right), \tag{9}$$

where $Y^{(t)}$ is the solution for iteration t, η represents the learning rate, and $\alpha(t)$ denotes the momentum for iteration t. The random normal distribution of the initial value $Y^{(0)}$ is usually set to $N(0, 10^{-4}I)$.

2.4. Implementation Steps. When the t-SNE algorithm is adopted to reduce high-dimensional data, if the dimension of data points is too large, then the algorithm will take a long time. In order to improve the efficiency of t-SNE, the PCA method is usually introduced first to reduce the dimension of a high-dimensional sample point data set to 50 dimensions, and then, t-SNE is used for cluster recognition. The specific pseudocode is shown in Algorithm 1.

3. Experiment and Analysis

3.1. Types of Samples. In this paper, two kinds of ceramic matrix composites (CMCs) and a silicon slice were selected as the detection objects for terahertz spectral imaging. The samples are 1 piece of alumina (Al_2O_3) ceramic sheet, 2 pieces of beryllium oxide (BeO) ceramic sheets, and 1 piece of monocrystalline silicon. For convenient comparison and analysis, the sample sheet is prepared for defect treatment. The specific specifications and defects are shown in Table 1.

3.2. Terahertz Spectrum of Samples. Nondestructive testing (NDT) method of spectral imaging was used to image the samples. The transmission time-domain spectrum of the sample is shown in Figure 2.

It can be seen from Figure 2(a) that the spectrum of alumina crack defect is significantly different from the normal spectrum, and the electric field intensity value is smaller than the peak intensity of the normal spectrum; also, the time delay is smaller than the normal spectrum. In Figure 2(b), the beryllium oxide defect part scatters a large amount of terahertz wave, resulting in severe attenuation of terahertz waves. The peak of the field strength of the differential spectrum is smaller than that of the normal spectrum, and the time delay is slightly smaller than that of the normal spectrum. In Figure 2(c), the spectrum of the monocrystalline silicon and the spectrum of the background reference signal have obvious differences in field strength and time delay.

3.3. Model Discriminant Analysis. The time-domain spectral data for each scan pixel of the samples are obtained directly by two-dimensional spectral scanning. To simplify data analysis, the original terahertz time-domain spectrum was directly used in this paper to establish the spectral data set of samples, and t-SNE was used to obtain the sample spectral data set. The scanning background spectrum was also included in the differential spectrum for cluster analysis.

Due to the high number of sample points and the spectral dimension, a random sampling method was adopted for spectral data set to reduce the time and complexity of model calculation. A certain number of sample points were randomly selected for model discrimination, and spectral clustering effect of the model was investigated

Algorithm: t-SNE
Input data: the sample terahertz spectral data set $X = \{x_1, x_2, \ldots, x_n\}$
Cost function: $C = KL(P\|Q) = \sum_i \sum_j p_{ij} \log_2 p_{ij}/q_{ij}$
Output the result: low-dimensional spatial data representation $Y(t) = \{y_1, y_2, \ldots, y_n\}$
Optimize training process
begin
  Set iteration times T, learning rate η, and momentum $\alpha(t)$
Calculate the perplexity Perp and the conditional probability according to $p_{ij} = p_{j|i} + p_{i|j}/2n$
  Randomly initialize $Y(0) = \{y_1, y_2, \ldots, y_n\}$ with a normal N (0, 10–4, I) distribution
  For $t = 1$ to T, do
    Calculate q_{ij} in lower dimensions with formula (5)
    Compute gradient $\delta C/\delta y$ according to formula (8)
    Update $Y^{(t)} = Y^{(t-1)} + \eta \delta C/\delta y + \alpha(t)(Y^{(t-1)} - Y^{(t-2)})$
  end
end
return Y

Algorithm 1: The pseudocode of t-SNE algorithm.

Table 1: The sample information of CMC and silicon.

The sample name	Specification	Description
Aluminum oxide (Al_2O_3) sheet with defects	Length × width × thickness (mm): 101.0 × 101.0 × 3.0 mm Bulk density ≥ 3.65 g/cm^3 Surface roughness Ra ≥ 0.5 μm Principal component content ≥ 95%	Gray-white, defects are internal irregularities cracks
Beryllium oxide (BeO) sheet with defects	Length × width × thickness (mm): 50.8 × 50.8 × 1.5 Bulk density ≥ 2.85 g/cm^3 Surface roughness Ra ≥ 0.1 μm Principal component content ≥ 99%	Pure white, highly toxic powder, the defect is the shape of the regular set of holes 4 Outside diameter: 4.2–5.9 mm Inner diameter: 2.6–4.2 mm Hole depth: 0.8 mm
Beryllium oxide (BeO) sheet with zero defects	Length × width × thickness (mm): 50.8 × 50.8 × 1.5 Bulk density ≥ 2.85 g/cm^3 Surface roughness Ra ≥ 0.1 μm Principal component content ≥ 99%	Pure white, highly toxic powder, zero defects
Monocrystalline silicon with zero defects	Thickness: 450 ± 10 μm Diameter: 76.3 ± 0.2 mm N dopant (phosphorus) Crystal orientation:<111>	Crystal growth: CZ resistivity: 14–45 Ωcm zero defects

under different perplexities and iteration times. The specific model parameter settings are shown in Table 2, and the model clustering results are shown in Figures 3–6, respectively.

3.3.1. Discriminant Result of BeO Sample with Defects. Theoretically, the ceramic sheet should have a good normal spectrum clustering effect by t-SNE, and the difference spectrum includes four sets of holes defect at the spectrum and the background spectrum scanning. It can be seen from Figures 3(c) and 3(d) that, under the condition of the same number of sampling points and iterations, the confusion levels of 30 and 50 have a similar clustering effect. As the number of sampling points increases to 10000, as shown in Figures 3(e) and 3(f), the degree of clustering of samples improves significantly, while the number of discrete clusters tends to decrease, but the demarcation of each cluster does not improve significantly. When the number of iterations increases to 5000, as shown in Figure 3(g), the clustering boundary of samples is significantly improved, and the increased spatial distance between clusters can reflect the actual classification of sample points, indicating that increasing the number of iterations can improve the stability of the model under the condition of large samples.

Under the large number of iterations, if the degree of confusion is increased, as shown in Figures 3(h) and 3(i), the sample clustering degree will also be strengthened, and more sample points with similar spectra are clustered together. Although the space distance between clusters has decreased, there is still clear dividing line between the surface. Especially in Figure 3(i), the actual situation of the sample points

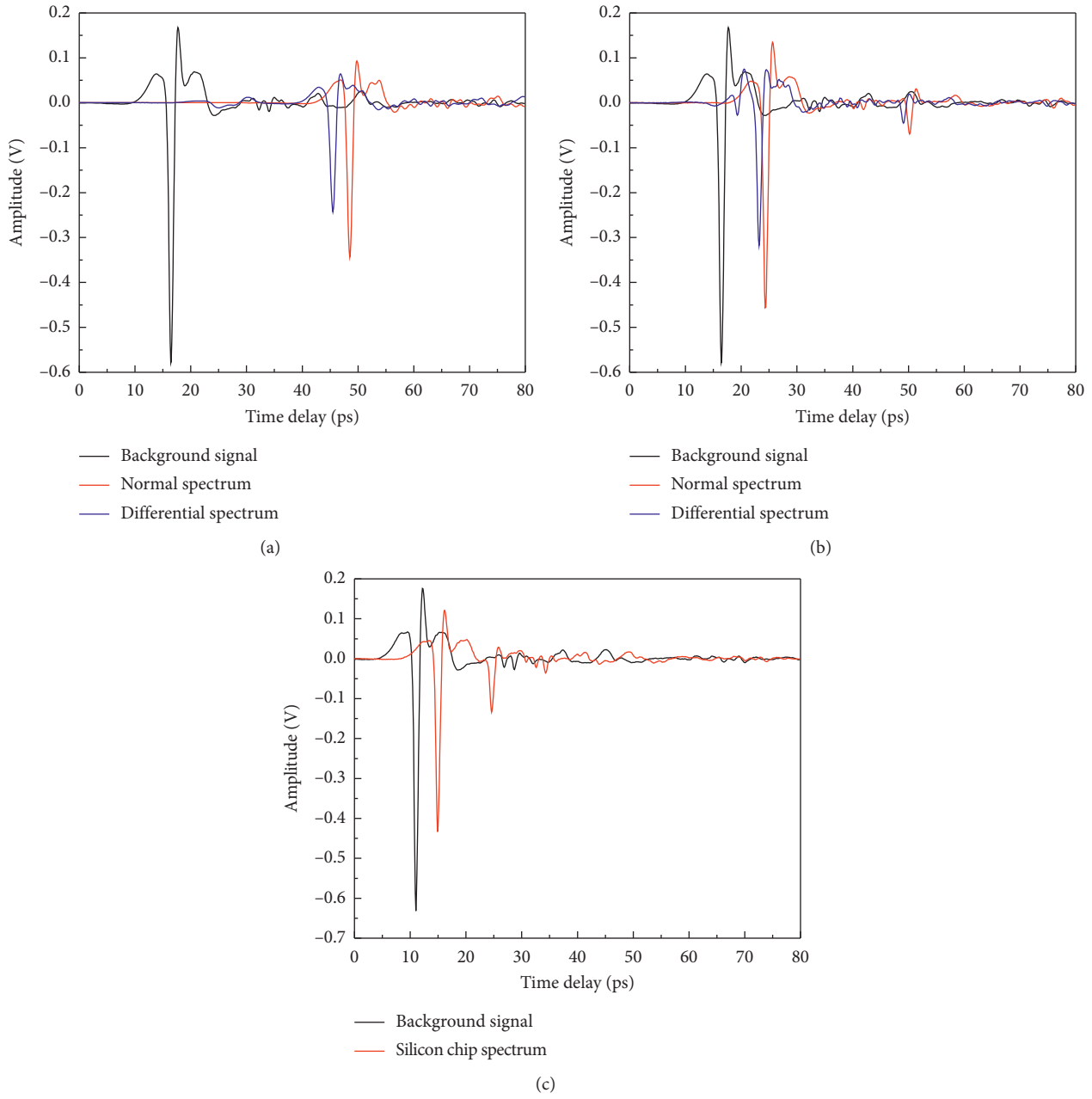


FIGURE 2: THz time-domain spectrum of sample imaging scanning: (a) transmission spectrum of Al_2O_3; (b) transmission spectrum of BeO; (c) spectrum of monocrystalline silicon.

classification is better reflected, so this picture can be chosen as t-SNE representative clustering view of the model. In general, the data set of abnormal spectrum is significantly smaller than that of normal structure samples and is in free state, and the results are in line with the predicted analysis.

3.3.2. Discriminant Result of the BeO Sample with Zero Defects. It can be seen from Figure 4 that, through observation and analysis on spectral images of BeO sample with zero defects, as shown in Figure 4(b), terahertz spectrum can be divided into two categories: background signal spectrum (air part) and normal BeO spectrum, and the number of reference signal spectra is smaller than that of BeO sample points. According to the cluster recognition result of difference spectrum, the experimental results are consistent with each other under different iteration times and sampling times when the perplexity is set to 100. The spectral data set was clearly clustered into two categories, and the classification boundaries were clear. In Figure 4(f), the classification boundaries of 5000

Table 2: Model parameters for four samples.

Sample name	Spectral dimension	PCA dimension reduction	t-SNE dimension	Maximum number of iterations	Perplexity	Number of spectra in data set	Number of sample points
BeO sheet with defects	800	50	2	1000, 5000	30–100	45240	2000–10000
BeO sheet with zero defects	800	50	2	1000, 5000	100	45240	3000–10000
Al_2O_3 sheet with defects	800	50	2	1000, 5000	100	8400	8400
Monocrystalline silicon	800	50	2	1000, 5000	100	108800	5000–12000

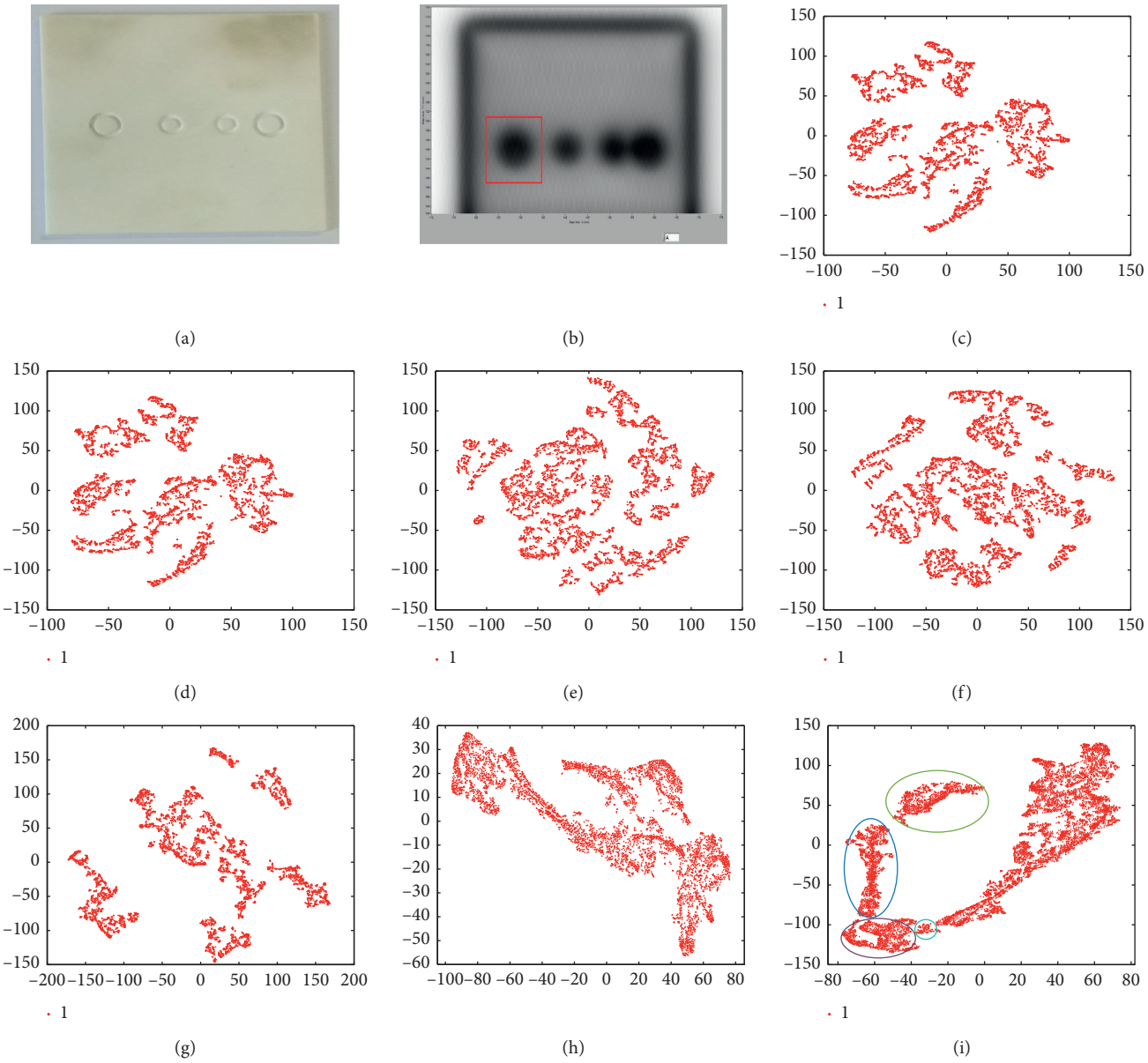


Figure 3: Clustering results of the BeO sample with defects: (a) optical image; (b) THz image; (c) 5000/1000/30; (d) 5000/1000/100; (e) 10000/1000/30; (f) 10000/1000/100; (g) 5000/5000/30; (h) 5000/5000/100; (i) 10000/5000/100. Description format: sampled data/iterations/perplexity.

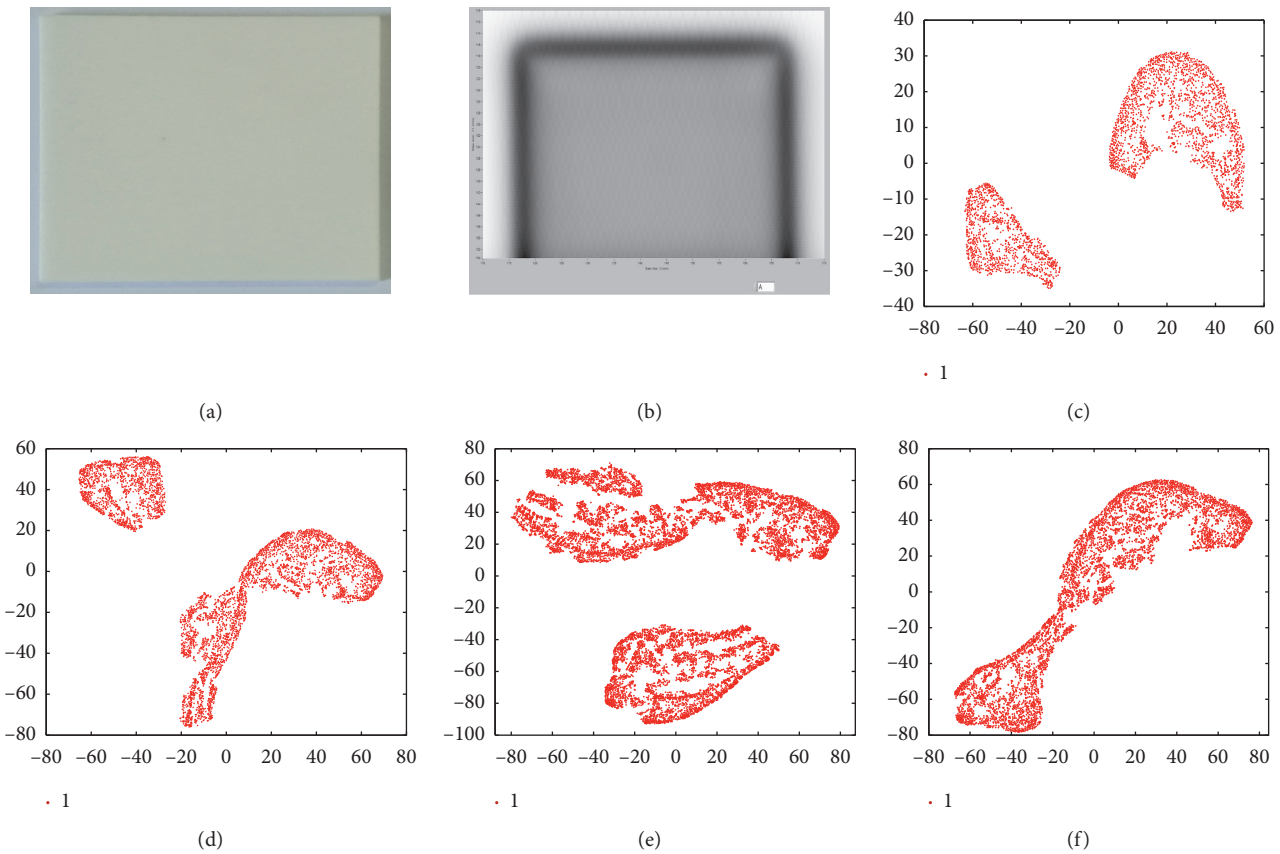


Figure 4: Clustering results of the BeO sample with zero defects: (a) optical image; (b) THz image; (c) 3000/1000/100; (d) 5000/1000/100; (e) 10000/1000/100; (f) 5000/5000/100. Description format: sampled data/iterations/perplexity.

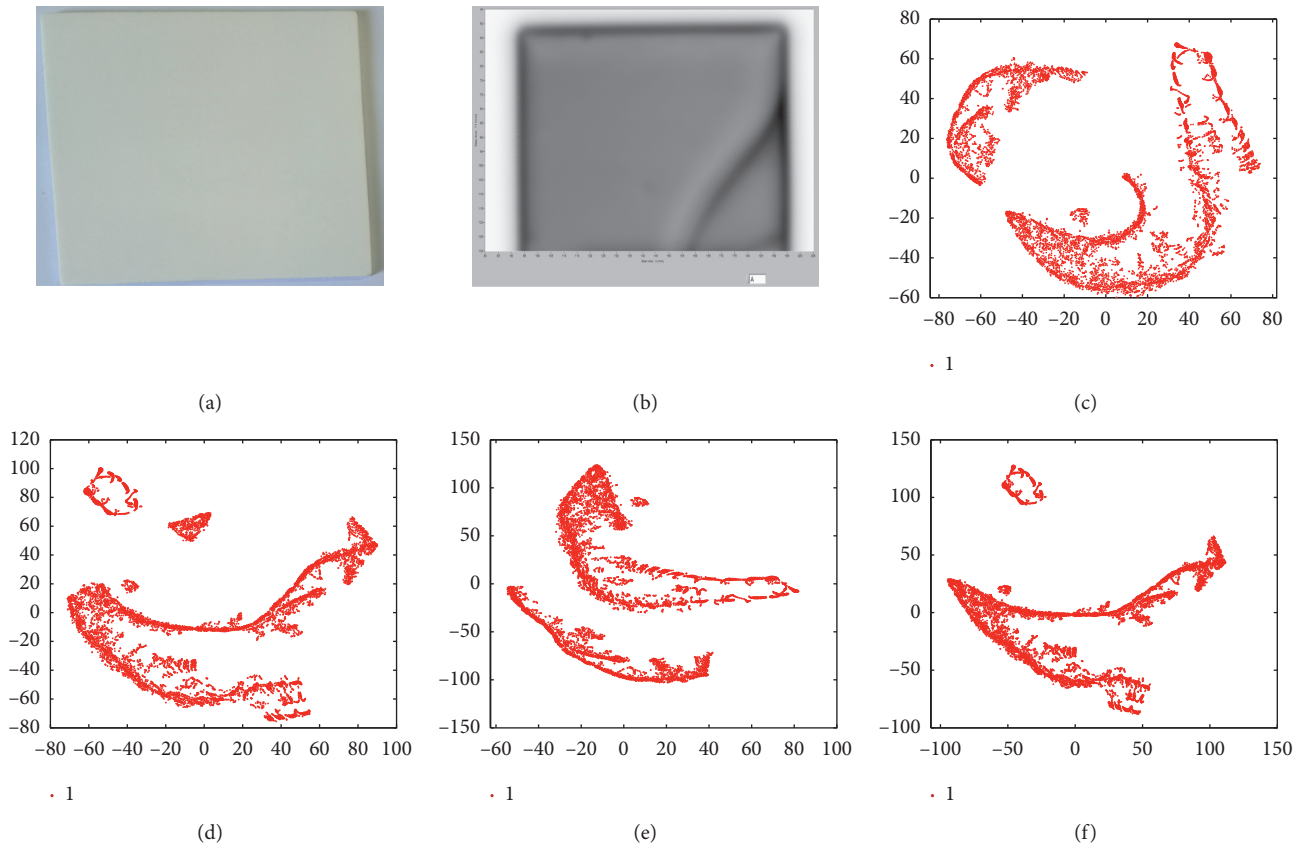


Figure 5: Clustering results of the Al_2O_3 sample with defects: (a) optical image; (b) THz image; (c) 8400/1000/100; (d) 8400/2000/100; (e) 8400/3000/100; (f) 8400/5000/100. Description format: sampled data/iterations/perplexity.

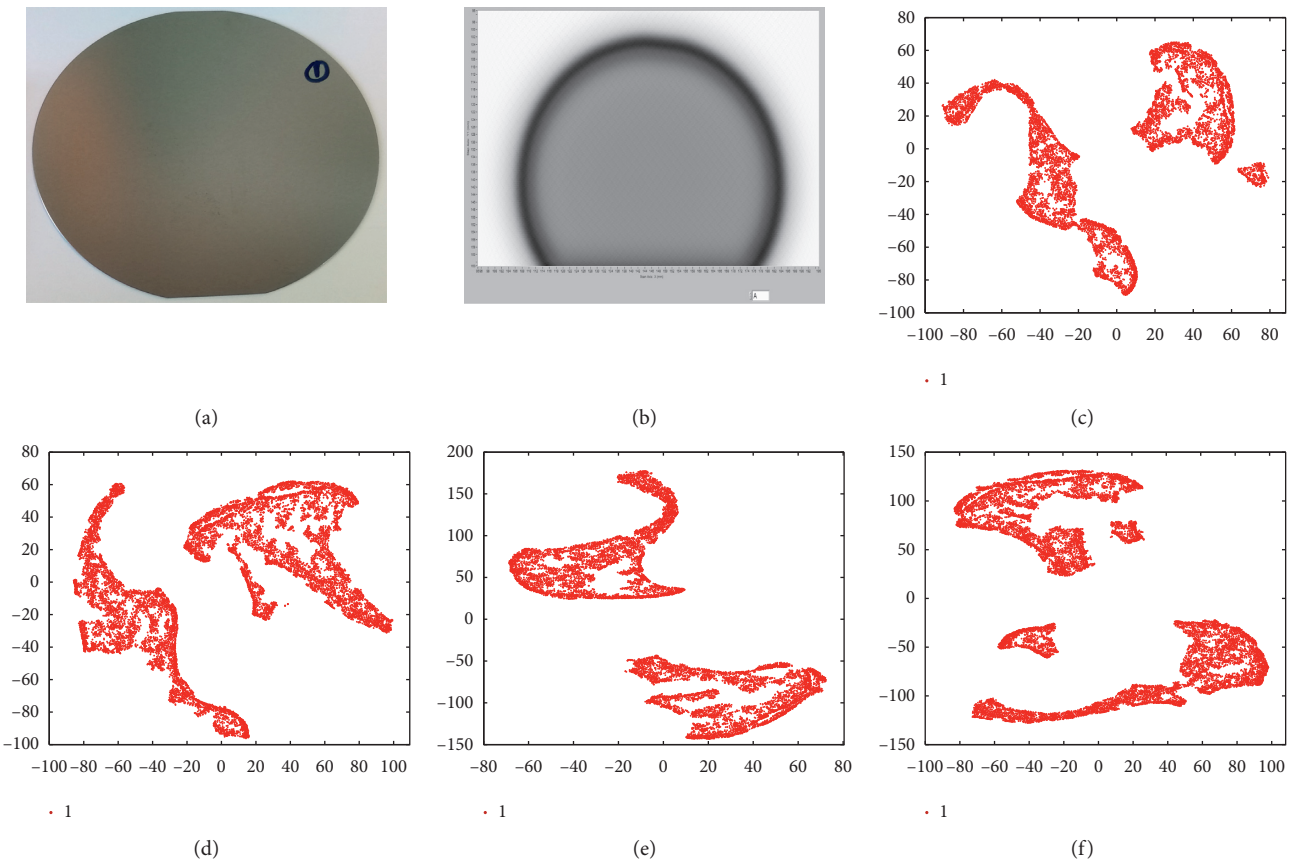


Figure 6: Clustering results of monocrystalline silicon with zero defects: (a) optical image; (b) THz image; (c) 8000/1000/100; (d) 12000/1000/100; (e) 10000/5000/100; (f) 12000/5000/100. Description format: sampled data/iterations/perplexity.

iterations were not as obvious as those of other iterations. In general, the reference spectral data set was significantly smaller than that of the beryllium oxide samples, which was in line with the predicted analysis results.

3.3.3. Discriminant Result of the Al_2O_3 Sample with Defects. As shown in Figure 5, cluster analysis is performed on the spectrum data set of the Al_2O_3 sample when the perplexity is set to 100. Under different iteration times, the experimental results are consistent, and the spectral data set is obviously clustered into four categories (the reference signal spectral set, the normal sample point spectral set, the background signal spectral set, and the sample point spectral set at the crack). Especially in Figure 5(d), the classification boundary of the spectral cluster is clear, and the scale of the differential spectral data set at the crack is much smaller than that of the normal spectral data set. The clustering results are consistent with the image features observed in the spectral image (Figure 5(b)). The results show that the t-SNE can be used for differential spectral clustering analysis to realize superresolution identification of sample defect features.

3.3.4. Discriminant Result of the Monocrystalline Silicon Sample. The clustering results of monocrystalline silicon chip are shown in Figure 6. As can be seen from the figure, the perplexity is set to 100, and consistent clustering results are obtained under different iteration times and sampling times. Spectral data sets are clearly clustered into two categories, and spectral clustering boundaries are quite clear. In Figures 6(c) and 6(f), small-scale discrete clusters can be clearly seen, but we are unable to observe obvious defects from the spectral image (Figure 6(b)), indicating that the sample may have a tiny defect structure lower than the optical resolution, which requires further imaging observation and analysis. This also reflects the effectiveness of t-SNE in superresolution identification of small defect structures.

4. Conclusion

For terahertz nondestructive testing, it is difficult to identify the structural features of samples with tiny defects due to the resolution limitation of the optical system. To solve this problem, we applied t-SNE to perform cluster analysis and identification of spectral data sets. Experiments on sample data set of scanned images indicate that t-SNE can precluster and identify the difference spectrum of the measured object. Furthermore, provide a priori predetection basis for the next step of pattern recognition classification and imaging analysis, thus improving the accuracy of spectral target recognition. In particular, the model clustering of the imaging spectral data set can overcome the inherent limitation

of optical resolution. From the perspective of spectral clustering, this method can provide a feasible method for realizing superresolution identification of samples which has an important research value for the rapid detection of large component samples in engineering applications.

The research method in this paper is different from the traditional method to identify target defects through images, and a new superresolution method to identify target defects through spectral clustering is created, which is an important auxiliary means to identify target defects through terahertz images. Since the method in this paper can only predict whether there is a defect in the detection target and the characteristics such as the type, shape, location, and size of the defect cannot be analyzed, the research focus in the next stage of this paper will aim to how to realize the judgment of defect characteristics through spectral recognition.

Data Availability

The data can be shared and used.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Support for Scientific Research Projects (Scientific Research Projects in Colleges and Universities) (nos. 440-99000332 and ZP2020038) and Foundation Funded Project of Doctoral (no. 440-99000617).

References

[1] D. H. Auston, K. P. Cheung, and P. R. Smith, "Picosecond photoconducting hertzian dipoles," *Applied Physics Letters*, vol. 45, no. 3, pp. 284–286, 1984.

[2] D. H. Auston, K. P. Cheung, J. A. Valdmanis, and D. A. Kleinman, "Cherenkov radiation from femtosecond optical pulses in electro-optic media," *Physical Review Letters*, vol. 53, no. 16, pp. 1555–1558, 1984.

[3] P. Lopato, G. Psuj, and B. Szymanik, "Nondestructive inspection of thin basalt fiber reinforced composites using combined terahertz imaging and infrared thermography," *Advances in Materials Science and Engineering*, vol. 2016, Article ID 1249625, 13 pages, 2016.

[4] L.-M. Xu, W.-H. Fan, and J. Liu, "High-resolution reconstruction for terahertz imaging," *Applied Optics*, vol. 53, no. 33, pp. 7891–7897, 2014.

[5] R. Heylen, M. Parente, and P. Gader, "A review of nonlinear hyperspectral unmixing methods," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 1844–1868, 2014.

[6] J. B. Tenenbaum, V. Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[7] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[8] G. E. Hinton, S. Geoffrey, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[9] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: a review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.

[10] V. D. L. Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.

[11] V. D. L. Maaten, "Accelerating t-SNE using tree-based algorithms," *Journal of Machine Learning Research*, vol. 15, pp. 3221–3245, 2014.

[12] G. E. Hinton and S. T. Roweis, "Stochastic neighbor embedding," *Advances in Neural Information Processing Systems*, vol. 15, no. 4, pp. 833–840, 2003.

[13] D. T. Montoro, F. Yuan, M. Biton et al., "A revised airway epithelial hierarchy includes CFTR-expressing ionocytes," *Nature*, vol. 560, no. 7718, pp. 319–324, 2018.

[14] D. A. Cusanovich, J. P. Reddington, D. A. Garfield et al., "The cis-regulatory dynamics of embryonic development at single-cell resolution," *Nature*, vol. 555, no. 7697, pp. 538–542, 2018.

[15] W. Li, J. E. Cerise, Y. Yang, and H. Han, "Application of t-SNE to human genetic data," *Journal of Bioinformatics and Computational Biology*, vol. 15, no. 4, Article ID 1750017, 2017.

[16] G. C. Linderman, M. Rachh, J. G. Hoskins, S. Steinerberger, and Y. Kluger, "Fast interpolation-based t-SNE for improved visualization of single-cell RNA-seq data," *Nature Methods*, vol. 16, no. 3, pp. 243–245, 2019.

[17] W. Jiang, J. Zhou, H. Liu, and Y. Shan, "A multi-step progressive fault diagnosis method for rolling element bearing based on energy entropy theory and hybrid ensemble auto-encoder," *ISA Transactions*, vol. 87, pp. 235–250, 2019.

[18] G. Traven, G. Matijevič, T. Zwitter et al., "The galah survey: classification and diagnostics with t-SNE reduction of spectral information," *The Astrophysical Journal Supplement Series*, vol. 228, no. 2, p. 24, 2017.

[19] E. Pouyet, N. Rohani, A. K. Katsaggelos, O. Cossairt, and M. Walton, "Innovative data reduction and visualization strategy for hyperspectral imaging datasets using t-SNE approach," *Pure and Applied Chemistry*, vol. 90, no. 3, pp. 493–506, 2018.

[20] W. Liu, P. Zhao, C. Wu, C. Liu, J. Yang, and L. Zheng, "Rapid determination of aflatoxin B1 concentration in soybean oil using terahertz spectroscopy with chemometric methods," *Food Chemistry*, vol. 293, pp. 213–219, 2019.

[21] G. Dimitriadis, J. P. Neto, and A. R. Kampff, "t-SNE visualization of large-scale neural recordings," *Neural Computation*, vol. 30, no. 7, pp. 1750–1774, 2018.

[22] E. Roman-Rangel and S. Marchand-Maillet, "Inductive t-SNE via deep learning to visualize multi-label images," *Engineering Applications of Artificial Intelligence*, vol. 81, pp. 336–345, 2019.

Hindawi
Discrete Dynamics in Nature and Society
Volume 2020, Article ID 4904829, 10 pages
https://doi.org/10.1155/2020/4904829

Research Article

Global Stability and Dynamic Analysis for a Type of Macroeconomic Systems

Ya-Juan Yang,[1,2] **Ru-Fei Ma**,[1] **and Jing Zhang**[3]

[1]*School of Business, Macau University of Science and Technology, Taipa, Macau*
[2]*City College, Dongguan University of Technology, Dongguan, Guangdong, China*
[3]*Department of Continuing Education, Dong Guan Polytechnic College, Dong Guan, China*

Correspondence should be addressed to Ya-Juan Yang; 1909853gbm30002@student.must.edu.mo

Received 26 June 2020; Accepted 29 July 2020; Published 28 August 2020

Guest Editor: Fuqiang Gu

This paper aims at the dynamic properties of the proposed globally planned economic systems named after CPE proposed by Loo-Keng Hua who is one of the worldwide famous Chinese mathematicians. First, we give new existence conditions of growth balanced solution to the model. Second, we lead into the concept of stability for balanced output and carry out a theorem that deals with some equivalent conditions for judging a solution of output starting from the fact that any initial input can whether approach the existing balanced solution or not. Third, a new dynamic price system related to interest factors is proposed here and it is demonstrated that the new price equation is a much generalized one in comparison with the original price one which is only a special case of this new price equation. Also, relationships of the balanced solutions between the price and the output equation are investigated and the stability analysis is studied as well for the new price system. Finally, two examples are employed to illustrate the technical operation of input-output method and some new contributions of this article.

1. Introduction

In the last century, undoubtedly, the most widely used economic theory is the input-output method which was proposed by Wassily Leontief, a winner of the Nobel Prize in economics in the year 1973. The seminal paper for the input-output method originated in 1953 by Leontief (see [1]) and the dynamic analysis for these types of economic systems was investigated in his two classical papers [2, 3]. A systematic discussion for input-output theory can be found in Leontief's book [4] and more references therein. Just as this method's validity in analyzing a country's economic development, there had been over 80 counties which prepared and reported their input-output economy statements by the input-output theory since 1979. Meanwhile, the United Nation's social and economic sections suggest its member that the input-output method should be led into the national statistics accounting systems.

Research of the input-output method for fully free competition market economy, FFCME in short, has received tremendous attention in the past six decades from the first proposal by Leontief and massive literature has been published since then. All these research articles, of the most important ones presented here, focused on various sorts of FFCME problems that could be typically classified into the following two types: one investigated the dynamical properties for Leontief's model (see [5–9]) and the other focused on the macroforecasting problems of the FFCME systems (see [10–13]). Status, progress, and application fields of Leontief's theory can be found in the latest symposium by Peterson (see [14]) while a complete and standard treatment to the theory here recommends to read the book by Miller and Blair [15].

Different from the FFCME system, there is another economic system named after centrally planned economy, CPE in short, characterized by central planning being chief while the market readjustment being accessorial and it has been executed in China over 60 years since the creation of PRC: exceeding 30 years, going on from 1949 to 1980, for complete CPE and 37 years for socialist market economy,

beginning from 1981. For the CPE system, Leontief accompanied President Nixon in an advisory capacity at the Beijing interview of President Nixon with Chairman Mao in 1987 and he then gave a high appraisal to the policy of the socialist market economy in his review article "socialism is practicable in China" when he came back to America. The input-output technology owns a wide application in the macroforecasting and planning of economic problems regardless of short term or long term as well as the ability to be applied to forecast and plan different economic policy systems, being whether the FFCME ones or CPE ones.

Compared with the massive research literature to the FFCME system, there has been much less research focusing on the CPE system until Professor Loo-Keng Hua, a world famous mathematician of China, starting with a seminal paper in USA [16], published a series of articles in Chinese Science Bulletin at the middle of 80s last century; see [17–23]. Of course, Professor Loo-Keng Hua's research extends Leontief's theory to the CPE system which is put into practice by China and hence is more suitable for analyzing the current economic development in China. Fundamentals of Professor Loo-Keng Hua's analysis are based on a hypothesis that China's productivity elasticity is large enough; it is allowable as Professor Blanchard points out in [24], so that the relationships between inputs and outputs can be then described well for the CPE system.

Although Professor Loo-Keng Hua established the mathematical models for analyzing the CPE system in articles as mentioned above and also some basic concepts and analysis, such as superior limit of production capacity, crisis of production system, price, and consuming coefficients, were introduced there; there are still many problems, especially dynamical properties, which need to be further investigated, for instance, the existence of this CPE system's balanced solution and its stability and price dynamics.

Motivated by this idea and in order to be in a parallel way in studying dynamical properties of the FFCME system [2, 3, 5–9], this paper first presents the existence of growth balanced solutions and their stabilities to the CPE system, and, then, a generalized dynamic price system is proposed with key merit that Loo-Keng Hua's primary price system is a special case of this. Meanwhile, relationships of the balanced solution between price and output are analyzed as well as the stability of the dynamic price system investigated too.

The remainder of this paper is organized as follows: Section 2 is an outline of CPE system formulation and some results of the study being made by Professor Loo-Keng Hua are reviewed there. The main study of this economic system in this paper is carried out in the next sections: new existence theorem for the growth balanced solution is addressed in Section 3 and the stability analysis for the output balance is in Section 4. Section 5 deals with the dynamic price system including the formulation and the stability of balanced price while Section 6 gives two examples for supporting our theoretic results. Finally, some concluding remarks are drawn in Section 7.

Remark 1. The so-called FFCME , also known as free enterprise economy, in which the production and sale of goods and services are completely guided by the free price mechanism of fully free competition market. Up to now, absolute fully free competition market economies do not exist, but their markets are open wider than others. The United States is generally regarded as the representative of the fully free competition market economy for it's higher market access. Typical examples of CPMs are the Soviet Union and China from 1949 to 1978. Actually, in today's world, many countries conduct "more than half" of fully free competition economies, such as today's China.

2. CPE Model and Leontief Model

In Leontief's macroeconomic model, the production activities of an economy are divided into n industrial sectors and product transactions between these sectors are analyzed based on the following basic assumptions:

H1: there is no joint production, and each industry sector produces only one product, which means there is a one-to-one correspondence between sectors and products, and so they can be substituted for each other.

H2: a single product produced by each sector requires a certain amount of input from other sectors; that is, sector j requires t_{ij} units of product i as input to produce a unit of the jth product; here $i = 1, 2, \ldots, n$ and t_{ij} are called the input coefficients.

H3: in the whole production process of an economy, there is no lag in production, no capital goods, no foreign trade, neither involvement of government activities.

Under the above assumptions, let x_i be the total output of the jth product and $x = (x_1, x_2, \ldots, x_n)^T$ the output vector. For production of the other n sectors, the amount consumed by x_i is

$$\sum_{j=1}^{n} t_{ij} x_i, \tag{1}$$

and the remaining amount will be

$$d_i = x_i - \sum_{j=1}^{n} t_{ij} x_i. \tag{2}$$

This net output d_i is called the final demand of the ith product, and let $d = (d_1, d_2, \ldots, d_n)^T$ represent the final demand vector.

Then, as the aggregate demand must be equal to the aggregate supply, we can get the following total input-output equation systems:

$$(I - T)x = d, \tag{3}$$

which describe the economy as a whole; here, $T = (t_{ij}) \in M_n(R)$ is called the input matrix and $L = I - T$ the Leontief matrix. These input-output systems (3) are called open Leontief model and they are the theoretical basis for analyzing the macroeconomic state of an economy.

Leontief's method gives an important contribution in the optimization of the entire production process since it is a very complex problem and there has not been a corresponding mathematical method which can fully solve this optimal task, especially the lack of mathematical description

for the large-scale economic behavior. The hypotheses (H1)(H3) of Leontief's model are very strict ones (only can be partially meted by FFCME as mentioned in Section 1), which are far from the actual economic operation state, but this does not affect the wide application of the input-output analysis for an economy.

Why do we say that? The most important economic activity in an economy is the production and the domestic sales and hence what we are interested in here are the economy's intersectoral input and the external demand, this may be the reason that foreign trade is excluded in hypothesis 3. Because Leontief's method just meets this purpose, the input-output reports have been employed by, as mentioned earlier, more than 80 countries in the world. While carrying out an economy's input-output report, there are many various input-output tables to be created and details for making an input-output table can be found in Section 6; see Example 1 therein.

Although the input-output analysis has become the standard template for much national economic planning, national development plans cover much more than economics such as population, welfare, and environment. For a country, an economy, taking these problems together, boils down to a multiobjective decision problem of which the first study is made by Schinnar who analyzes development planning in a Leontief input-output model in order to take into account both economic and demographic goals; see [25].

Thereafter, Chichilnisky carried out first the green golden rule in respect of social welfare and then defines social welfare as a weighted average between the discounted utilitarianism and the green golden rule welfare; see [26, 27]. Up to now, we believe that Leontief's model combined with the multigoal decision-making model will be the most promising research topics for addressing macroeconomic planning and forecasting, for which the interested readers may refer to the review article [28] by Cinzia et al.

Different from Leontief's input-output model, Professor Loo-Keng Hua proposed a CPE system model (see [17, 18]), where the national production is divided into n sectors also, which are interrelated with a so-called consumption coefficient matrix $A = (a_{ij})_{n\times n}$, with a_{ij} being the quantities of the jth class to be consumed for producing a unit quantity of the ith class. It is obvious that $A = (a_{ij})_{n\times n} \geq 0$ and $a_{ij} > 0$ for some i, j.

By letting X^t , normally being a n−dimensional column vector, be the output of the t−period, Loo-Keng Hua established the CPE system as follows:

$$(\text{CPE})\ X^t = AX^{t+1}, \quad t \in T = \{0, 1, 2, \ldots,\}. \tag{4}$$

If A is an invertible matrix, the previous equation, equivalently, can be formulated by

$$(\text{CPE})\ X^{t+1} = A^{-1}X^t, \quad t \in T = \{0, 1, 2, \ldots,\}. \tag{5}$$

This CPE system is based on the assumption that the production of every class grows in a fixed proportion at each period t and the consumption coefficient matrix does not vary with time t. It is easy to see, by the inductive method, that $X^{(t)} = (1/\lambda^t)X^{(0)}$ will be the $t-$ period production, namely, a solution of the CPE system (4) or (5) if the initial input starting from an eigenvector $X^{(0)} > 0$ of A with the corresponding eigenvalue λ.

It should be noted that if we let $t = 0$, by (4), then $X^0 = \mathrm{AX}^1$; substituting it to $X^1 - X^0$ gives us

$$X^1 - X^0 = (I - A)X^1, \tag{6}$$

which is the same as Leontief's model (3) with $d = X^1 - X^0$, $x = X^1$, and $T = A$. So, in this sense, Leontief's model is a special case of the proposed CPE model.

At the same time, Loo-Keng Hua also gave a price equation coupled with equation (4) as follows:

$$(q_1, q_2, \ldots, q_n)\lambda_* = (q_1, q_2, \ldots, q_n)A, \tag{7}$$

where q_i is the price per-unit of the ith class and λ_*, also called the "price changing rate," is the largest eigenvalue of A. According to Loo-Keng Hua's primary definition, the price defined here means the product's basic output value only without the profits being contained; see [20].

For the CPE system (4), Loo-Keng Hua investigated the fundamental dynamical property (see Theorem 1 of [18]), where it is called "basic theorem" and we keep it here.

Basic Theorem: let A be a nonnegative irreducible square matrix, for any noneigenvector of A but positive vector $x > 0$; there exists a positive integer l_0 such that

$$xA^{-l}, \tag{8}$$

must be variable vector when $l \geq l_0$, namely, some entries of xA^{-l} are positive and some negative.

It follows from this Basic Theorem that some outputs of production will be negative in several years if the initial input is a positive noneigenvector of A and, thus, the economic system will collapse. So, the Basic Theorem reveals a basic principle of philosophy that balance is temporary but imbalance is perpetual for any economic system: imbalance will ultimately happen and even collapse, someday if the initial input x is not an eigenvector of the consumption coefficient matrix A.

Of course, it is far from enough to stop at this Basic Theorem for the dynamical properties of the CPE system (4) as a much deeper study with respect to the dynamical property for the FFCME system is investigated by [2, 3, 5, 25].

To proceed, we need some concepts of dynamical property for economic systems such as definitions of the balanced solution and the causal indeterminacy proposed by Solow and Samuelson in [29].

Definition 1. For the CPE system (4), if there exists a constant $\alpha > 0$ such that $X^t = \alpha X^{t-1}$ for all $t \in N = \{1, 2, \cdots\}$, then X^t is called a balanced solution of (4) with the growth rate α.

It is clear that if the consumption coefficient matrix A is a nonnegative irreducible square matrix, then, by the well-known Perron–Frobenius theorem, there exists a positive eigenvector X of A that must be a balanced solution of the CPE system (4) with the growth rate $(1/\lambda_*)$; here λ_* is the largest positive eigenvalue of A and X, being the

corresponding eigenvector. Obviously, by Definition 1, the economic quantity of the CPE system (4) will increase as $(1/\lambda_* > 1)$ and decrease as $(0 < 1/\lambda_* < 1)$.

The case occurring in the Basic Theorem is called the causal indeterminacy for an economic system as defined in the following definition.

Definition 2. If the initial input of an economic system does not meet any possible balanced solution, then there must exist at least one output to be negative. The economic system that owns this property will be known as having the causal indeterminacy.

Remark 2. By the Basic Theorem, it follows that the causal indeterminacy must happen for the CPE system (4) if the consumption coefficient matrix A is a nonnegative irreducible square matrix.

3. Existence of the Growth Solution

As stated in Remark 2 of the previous section, the CPE system (4) yields a balanced solution with the economic growth rate $(1/\lambda_*)$, if the consumption coefficient matrix A is a nonnegative irreducible square matrix, thus the economic system will develop well when $(1/\lambda_* > 1)$ since the economic growth keeps on. It is certain that there are some other existence conditions on the balanced growth solution of the CPE system (4) except for the ones proposed by Loo-Keng Hua and it would be best to find these conditions for the existence of the growth balanced solution so as to control the economic system developing under the way. Meanwhile, this section will focus on these themes then.

Different from Loo-Keng Hua's assumption on the consumption coefficient matrix A being an irreducible matrix, here we try to find the existence condition for the more general matrix ones. First, we need two lemmas for proving the main existence theorem.

Lemma 1. *Let $A = (a_{ij})_{n\times n}$ be a nonnegative real matrix with a nonnegative real eigenvalue $\lambda \geq 0$ that dominates all the other eigenvalues $\lambda_i\,(i = 1, 2, \ldots)$ of A in absolute value, that is, $\lambda \geq |\lambda_i|$, $i = 1, 2, \ldots$; then, for any positive vector $x = (x_1, x_2, \ldots, x_n)^T$, we have*

$$\min_{1\leq i\leq n}\left(\frac{1}{x_j}\sum_{j=1}^{n} a_{ij}x_j\right) \leq \lambda \leq \max_{1\leq i\leq n}\left(\frac{1}{x_j}\sum_{j=1}^{n} a_{ij}x_j\right). \tag{9}$$

Proof. See [30] Theorem 8.1.26. □

Lemma 2. *Suppose A is the consumption coefficient matrix of CPE system (4) and $I - A$ is a diagonal strictly dominant matrix; then there exist a positive real number μ: $0 < \widetilde{\mu} < 1$ and a positive vector $\widetilde{x} \geq 0$ such that $A\widetilde{x} = \widetilde{\mu}\widetilde{x}$.*

Proof. Since A is the consumption coefficient matrix of CPE system (4), we have $A \geq 0$. Then, it follows by the well-known Perron–Frobenius theorem that there exist a positive real number $0 \leq \widetilde{\mu} \leq 1$ and a positive vector $\widetilde{x} \geq 0$ such that $A\widetilde{x} = \widetilde{\mu}\widetilde{x}$.

By Gershgorin theorem, all the eigenvalues of A belong to the set

$$\bigcup_i \Omega_i = \left\{\mu \middle| \mu - a_{ii}\right| \leq \sum_{j\neq i} a_{ij}\Bigg\}. \tag{10}$$

Thus, we have $|\widetilde{\mu} - ta_{ii}| \leq \sum_{j\neq i} a_{ij}$.

By this and the assumption of $I - A$ being a diagonal strictly dominant matrix, that is, $1 - a_{ii} > \sum_{j\neq i} a_{ij}$, we get $2a_{ii} - 1 < \widetilde{\mu} < 1$ and hence $\widetilde{\mu} < 1$.

In addition, according to Lemma 1, by letting $x = (x_1, x_2, \ldots, x_n)^T = (1, 1, \ldots, 1)^T$, we get

$$\min_{1\leq i\leq n}\left(\sum_{j=1}^{n} a_{ij}\right) \leq \widetilde{\mu} \leq \max_{1\leq i\leq n}\left(\sum_{j=1}^{n} a_{ij}\right). \tag{11}$$

So, according to the nonnegative of A, namely, $\sum_{j=1}^{n} a_{ij} \geq 0$, as well as the economic meaning of A, we get that $\sum_{j=1}^{n} a_{ij} \geq 0$ and hence $\widetilde{\mu} > 0$. Therefore, it follows that there exist $\widetilde{\mu}$: $0 < \widetilde{\mu} < 1$ and a vector $\widetilde{x} \geq 0$ such that $A\widetilde{x} = \widetilde{\mu}\widetilde{x}$. □

Now, we can give the main result about the existence of growth balanced solution for the CPE system (4).

Theorem 1. *Let $A = (a_{ij})_{n\times n}$ be the consumption coefficient matrix of CPE system (4) with the assumption of $1 - a_{ii} > \sum_{j\neq i} a_{ij}$; it could be satisfied when $I - A$ is a diagonal strictly dominant matrix. Then, there exists a growth balanced solution $X^{(t)} = (1/\widetilde{\mu})^t\widetilde{x}$ of system (4) with the growth rate $(1/\widetilde{\mu})$; here $\widetilde{\mu}$: $0 < \widetilde{\mu} < 1$ is the eigenvalue of A and $\widetilde{x} \geq 0$ the corresponding eigenvector.*

Proof. First, letting $t = 0$, by system (4), we get

$$X^0 = AX^1. \tag{12}$$

If X^0 is a balanced solution of (12) with the growth rate $(1/\lambda)$, then by Definition 1, we get

$$X^1 = \frac{1}{\lambda}X^0. \tag{13}$$

It follows from (12) and (13) that

$$\lambda X^0 = \mathrm{A}X^0. \tag{14}$$

Inductively, it is easy to get that $X^t = (1/\lambda)^t X^0$ will also be a balanced solution.

Second, by letting $X^0 = \widetilde{x}$ and according to Lemma 2, we obtain that $X^{(t)} = (1/\widetilde{\mu})^t\widetilde{x}$ will be a growth balanced solution with growth rate $(1/\widetilde{\mu}) > 1$. Therefore, Theorem 1 is proved then. □

Remark 3. Theorem 1 gets rid of the restriction that the consumption coefficient matrix should be irreducible. Further, under the condition proposed here, it is impossible that the total costs in quantities for each class will be more than the total outputs to be produced. This is the reason for the occurrence of negative output then. So, controlling the

economic system (4) in this way will be a better choice since there is no causal indeterminacy occurring under this direction for the CPE system (4).

4. Stability of the Balanced Output

As stated in the previous section, a better way to control an economic system is to avoid the causal indeterminacy occurring and find the growth solution so as to keep the economy growing. Though Theorem 1 gives a good description of finding out the growth balanced solution, the condition for this needs a rigid initial point so as to reach the goal.

If started at any point, it is vitally important to investigate whether this solution can approach the existing growth balanced solution or not, namely, the stability of the economy balance. First, we need a mathematical description for the stability; see the following definition.

Definition 3. Let $A \geq 0$ be the consumption coefficient matrix of system (4) and assume it is an invertible matrix. Suppose $X^{*(t)} = (1/\mu^t)X > 0$ is a balanced solution of system (4); if for every solution $\widehat{X}^{(t)}$ starting from any initial input $\widehat{X}^{(0)} \geq 0$, there exists a constant σ: $0 < \sigma < \infty$ such that $\lim_{t\longrightarrow\infty} (\widehat{x}_i^{(t)}/x_i^{*(t)}) = \sigma$, where $\widehat{x}_i^{(t)}, x_i^{*(t)}$ is the ith entry of $\widehat{X}^{(t)}$ and $X^{*(t)}$, respectively, then the balanced solution $X^{*(t)}$ is called a stable balanced solution.

Clearly, if the CPE system (4) yields a stable balanced solution, then all the solutions determined by any initial input will eventually be greater than zero since any of these solutions approaches asymptotically to the positive balanced solution $X^{*(t)} > 0$. Thus, there has been no causal indeterminacy happening in this case for system (4).

Next, we give a necessary and sufficient condition to the stability of a balanced solution for system (4).

Theorem 2. *Suppose $A \geq 0$ is the consumption coefficient matrix of system (4) as well as an invertible matrix; then we have the following:*

(i) There exists a positive integer t such that $(A^{-1})^t > 0$ if and only if $|\lambda_i| > \lambda_1 > 0, X_1 > 0$, where $\lambda_i, i = 1, 2, \ldots, n$ are the eigenvalues of A and X_i the corresponding eigenvectors.

(ii) The balanced solution $X^{*(t)} = (1/\lambda_1^t)X_1 > 0$ is stable if and only if $|\lambda_i| > \lambda_1 > 0$, where $\lambda_i, i = 1, 2, \ldots, n$ are the eigenvalues of A and X_i the corresponding eigenvectors.

Proof. First of all, let us prove that (i) for any invertible matrix $A \geq 0$, there exists a positive integer t such that $(A^{-1})^t > 0$ if and only if $|\lambda_i| > \lambda_1 > 0, X_1 > 0$, where $\lambda_i, i = 1, 2, \ldots, n$ are the eigenvalues of A and X_i the corresponding eigenvectors.

If $(A^{-1})^t > 0$ for some positive t, then $(A^{-1})^t > 0$ will be an irreducible matrix. It follows by the well-known Frobenius theorem that matrix $((A^{-1})^t)'s$ eigenvalues μ_i and the corresponding eigenvectors $Y_i (i = 1, 2, \ldots, n)$ must satisfy $\mu_1 > |\mu_i|, Y_1 > 0$ with $AY_1 = \mu_1 Y_1$, where, $i = 2, 3, \ldots, n$. Let λ_i be the eigenvalues of A and X_i the corresponding eigenvectors; then it is easy to say $\mu_i = (1/\lambda_i^t)$, $i = 1, 2, \ldots, n, Y_1 = X_1$. So, by $\mu_1 > |\mu_i|$, we get $(1/\lambda_1^t) > |(1/\lambda_i^t)|$ $(i = 2, 3, \ldots, n)$ and hence $|\lambda_i| > \lambda_1 > 0$, $X_1 > 0$.

On the contrary, let λ_i be the eigenvalues of A and X_i the corresponding eigenvectors with $|\lambda_i| > \lambda_1 > 0, X_1 > 0$, where $i = 2, 3, \ldots, n$. First, any vector V could be represented as a linear combination of X_i for the invertibility of X_i; that is,

$$V = h_1 X_1 + h_2 X_2 + \cdots + h_n X_n. \tag{15}$$

Let $V = E^i$ be the column vector of the unit matrix E, that is, all the entries of E^i being 0 except for the ith being 1; then we have

$$E^i = h_1 X_1 + h_2 X_2 + \cdots + h_n X_n. \tag{16}$$

Making the inner product of this E^i with X_1, we get

$$\left[X_1, E^i\right] = \left[X_1, h_1 X_1 + h_2 X_2 + \cdots + h_n X_n\right] = \sum_{i=1}^{n} h_i \left[X_1, X_i\right]. \tag{17}$$

Note that $[X_1, X_i] = 0$ for $i \neq 1$; we have $[X_1, E^i] = h_1 [X_1, X_1]$ and so $h_1 = ([X_1, E^i]/[X_1, X_1])$. h_1 must be positive since $X_1 > 0$ and the choice of E^i.

Similarly, any solution $\widehat{X}^{(t)}$ with starting initial input E^i for system (5) can be reformulated as

$$\widehat{X}^{(t)} = \left(A^{-1}\right)^t E^i = h_1 \frac{1}{\lambda_1^t} X_1 + h_2 \frac{1}{\lambda_2^t} X_2 + \cdots + h_n \frac{1}{\lambda_n^t} X_n, \tag{18}$$

or, equivalently,

$$\lambda_1^t \widehat{X}^{(t)} = \lambda_1^t \left(A^{-1}\right)^t E^i = h_1 X_1 + h_2 \frac{\lambda_1^t}{\lambda_2^t} X_2 + \cdots + h_n \frac{\lambda_1^t}{\lambda_n^t} X_n. \tag{19}$$

By $|\lambda_i| > \lambda_1 > 0$, that is, $0 < (\lambda_1/|\lambda_i|) < 1$, and paying attention to $X_1 > 0, h_1 > 0$, we get that, for sufficient enough, large $t \geq N$,

$$\lambda_1^t \widehat{X}^{(t)} = \lambda_1^t \left(A^{-1}\right)^t E^i > 0. \tag{20}$$

As $\lambda_1^t > 0$, we obtain $(A^{-1})^t E^i > 0$ for large t. Since it holds for any E^i, we get that $(A^{-1})^t > 0$.

Secondly, we prove (ii). If $|\lambda_i| > \lambda_1 > 0$, according to Definition 3 and (18) we have

$$\lim_{t\longrightarrow\infty} \frac{\widehat{x}_i^{(t)}}{x_i^{*(t)}} = \frac{h_1 \left(1/\lambda_1^t\right) X_{1i} + h_2 \left(1/\lambda_2^t\right) X_{2i} + \cdots + h_n \left(1/\lambda_n^t\right) X_{ni}}{\left(1/\lambda_1^t\right) X_{1i}}$$

$$= h_1. \tag{21}$$

Let $\sigma = h_1$ and note that $h_1 > 0$; we get that this $X^{*(t)}$ should be a stable balanced solution.

On the other hand, if $\lim_{t\longrightarrow\infty} (\widehat{x}_i^{(t)}/x_i^{*(t)}) = \sigma > 0$, then, for sufficiently large $t \geq N$, we have $\widehat{x}_i^{(t)} > 0, i = 1, 2, \ldots, n$. So, $\widehat{X}^{(t)} > 0$ for sufficiently large $t \geq N$ with any initial input vector $\widehat{X}^0 \geq 0$. Similar to the previous proof of (i), if we let

$\widehat{X}^0 = E^i$, then we obtain $\widehat{X}^{(t)} = (A^{-1})^t E^i > 0$ for all $i = 1, 2, \ldots, n$ and sufficiently large $t \geq N$. Thus, $(A^{-1})^t > 0$. Therefore, by proving (i), we get that $|\lambda_i| > \lambda_1 > 0, X_1 > 0$. □

Remark 4. Up to now, for the CPE system (4), under the assumption of the consumption coefficient matrix $A \geq 0$ being invertible, we have the following equivalent results: the balanced solution $X^{*(t)} = (1/\lambda_1^t)X_1 > 0$ is stable ⇔ there exists a positive integer t such that $(A^{-1})^t > 0 \Leftrightarrow |\lambda_i| > \lambda_1 > 0, X_1 > 0 \Leftrightarrow$ causal indeterminacy cannot occur ⇔ each product's output cannot be negative in the subsequent years later.

5. Dynamic Price System

Loo-Keng Hua's price equation (7) is based on not considering the capital costs such as the money rate. At the same time, it gives an assumption that the product's price changes in a fixed proportion, for instance, λ^*; that is $P^{(t+1)} = \lambda_* P^{(t)}$ for each period t. But, certainly, interest rate impacts directly on the product's sale price while the product's sale price may generally not change in a fixed proportion in practical economic environment. Hence, we need to rebuild or extend Loo-Keng Hua's price system so as to match the practical economic system soundly and this will be carried out in the following subsection 5.1.

In addition, subsection 5.2 deals with the dynamical properties for the proposed price system such as the balanced price solution and its stability as well. The deep research on this respect will contribute to controlling the economic system (4) in a way of keeping the economy growing continually.

5.1. Formulation of the Price System. Let $P_i^{(t)}$ be the price of the ith class per-unit product in the t-period. The price vector for all n class products in the t-period could be represented as $P^{(t)} = (P_1^{(t)}, P_2^{(t)}, \ldots, P_n^{(t)})$. Assume any product's price to be a fixed constant during each period and costs of other products consumed for producing any product to be paid at the beginning of each period. Then, the per-unit cost for producing the jth class product in the t-period is

$$v_j^{(t)} = \sum_{i=1}^{n} P_i^{(t)} a_{ij} = P^{(t)} a_j, \tag{22}$$

where a_j is the jth column of the consumption coefficient matrix A.

The per-unit profit, regardless of other costs, for producing the jth class product in the t-period will be

$$\pi_j^{(t)} = P_j^{(t+1)} - P^{(t)} a_j. \tag{23}$$

At the same time, if money to be loaned out can get an interest rate r^t during the time between t-period and $(t+1)$-period, then the capital revenue for buying $v_j^{(t)}$ will be

$$R_j^{(t)} = r^t v_j^{(t)} = r^t P^{(t)} a_j. \tag{24}$$

By the competition arbitrage principle, interest and profit should eventually reach an equilibrium state; that is,

$$\pi_j^{(t)} = R_j^{(t)}. \tag{25}$$

So,

$$P_j^{(t+1)} - P^{(t)} a_j = r^t P^{(t)} a_j, \tag{26}$$

or

$$P_j^{(t+1)} = \left(1 + r^t\right) P^{(t)} a_j. \tag{27}$$

As (27) holds for all $j = 1, 2, \ldots, n$, we can formulate it in the following matrix representation:

$$P^{(t+1)} = \left(1 + r^t\right) P^{(t)} A. \tag{28}$$

Letting $M = (1 + r^t)A$, we obtain the dynamic price equation corresponding to (4) as follows:

$$P^{(t+1)} = P^{(t)} M. \tag{29}$$

If interest rate $r^t = 0$ and the price changing rate varies with a fixed proportion $\lambda_* = (P_i^{t+1}/P_i^t)$ for each period t, then this dynamic price equation (29) becomes the price equation (7) proposed by Loo-Keng Hua. Next, we turn to investigate the balanced price solution and the stability for the price equation (29) as made previously for the output equation (4).

5.2. Stability of the Balanced Price. In order to simplify, we suppose the interest rate $r^t = r$ to be a constant and $A \geq 0$ to be invertible. It is easy to see that $M = (1 + r)A$ is also invertible and the eigenvalues of M will be $(1 + r)\lambda_i$ if λ_i is the eigenvalues of A while the corresponding eigenvectors are the same.

Definition of the balanced price solution to (29) can be defined as Definition 1 for output equation (4); that is, if $P^{(t)} = \beta P^{(t-1)}$ holds for all $t \in N = \{1, 2, \ldots, n, \ldots\}$ and some constant $\beta > 0$, then $P^{(t)}$ will be a balanced solution of (29) with price changing rate β. The definition of stability for a balanced price solution can be found in the following mathematical description.

Definition 4. Let $A \geq 0$ be an invertible matrix and $M = (1 + r)A$. Suppose $P^{*(t)} = \zeta_1^t p_1 > 0$ is a balanced price solution of system (29); if, for a solution $\widehat{P}^{(t)}$ starting from any initial price $\widehat{P}^{(0)} \geq 0$, there exists a constant σ: $0 < \sigma < \infty$ such that $\lim_{t \longrightarrow \infty} (\widehat{p}_i^{(t)}/p_i^{*(t)}) = \sigma$, where Z is the ith entry of $\widehat{P}^{(t)}$ and $P^{*(t)}$, respectively, then the balanced solution $P^{*(t)}$ is called a stable price solution.

Similarly, we have Theorem 3 which reveals the dynamical properties for price equation (29) like Theorem 2 for the output equation (4).

Theorem 3. *Suppose $A \geq 0$ to be an invertible matrix and let $M = (1 + r)A$, where $r > 0$ is the interest rate. Then, we have the following:*

(i) There exists a positive integer t such that $M^t > 0$ if and only if $\mu_1 > |\mu_i| \geq 0, p_1 > 0$, where μ_i and p_i, $i = 1, 2, \ldots, n$, are the eigenvalues of M and the corresponding eigenvectors.

(ii) The balanced price $P^{*(t)} = \zeta_1^t p_1 > 0$ is stable if and only if $\zeta_1 > |\zeta_i| \geq 0, p_1 > 0$, where ζ_i and p_i, $i = 1, 2, \ldots, n$, are the eigenvalues of M and the corresponding eigenvectors.

Proof. The proof of this theorem is similar to Theorem 2 and a survey is given here.

First, it is noted that $A \geq 0$ and being invertible implies $M = (1 + r)A \geq 0$ and being invertible also. Second, if μ is an eigenvalue of matrix M with eigenvector p, then μ^t will be the eigenvalue of M^t with the same eigenvector p in regardless of scalar times. Third, inductively, it is easy to obtain $P^{(t)} = P^{(0)}M^t$. Finally, note that the general solution of equation (29) can be written as

$$\widehat{P}^{(t)} = \alpha_1 \zeta_1^t p_1 + \alpha_2 \zeta_2^t p_2 + \cdots + \alpha_n \zeta_n^t p_n, \tag{30}$$

where α_i are constants to be determined by the initial price $\widehat{P}^{(0)}$. All these four points give the key respects to prove Theorem 3 and the procedure of the proof is omitted here. □

Remark 5. The solutions $X^{*(t)} = (1/\lambda_1^t)X_1$ and $P^{*(t)} = \zeta_1^t p_1$ are a pair of output balance and price balance by the correspondence $\zeta_1 = (1 + r)\lambda_1, X_1 = p_1$. In fact, eigenvalues λ_i of A and eigenvalues ζ_i of $M = (1 + r)A$ have a correspondence $\zeta_i = (1 + r)\lambda_i$. So, if $\zeta_1 > |\zeta_i|$, then $\lambda_1 > |\lambda_i|$ and if $|\lambda_i| > \lambda_1$, then $|\zeta_i| > \zeta_1$. Thus, by Theorem 2 and Theorem 3, balanced price solution $P^{*(t)}$ being stable means balanced output $X^{*(t)}$ is not stable and vice versa.

6. Illustrative Examples

In this section, we illustrate some of the main results of this paper with two examples: one is about the input-output table formulation and the other is an application of Theorem 1 as well as a comparison declaration. First, let us give a simplified example to show the process of how to create an input-output table.

Example 1. Assuming a hypothetical economy is composed of (1) agriculture, (2) the industrial sector (manufacturing), and (3) the service provider (service) , each of these departments produces only one type of product, namely, the agricultural, industrial, or service supply, and there is interdependence between and among them. Each department buys products from the other departments and sells its own products to the opposite ones, but the final product and service supply (they do not enter the production process) is used by external departments, such as consumers. The production process and the external demand are not associated with the government and there is no foreign trade; then, according to the early hypotheses (H1)(H3), we can carry out a form to summarize the product and the current situation of service supply, as shown in Table 1, where x_{ij} represents products (in US $) sold by sector i to sector j.

The data of each row in Table 1 represents the allocation of the total output to different departments and users, while the data of each column represents the sources of department inputs required for the total output. For example, the first row shows that the total output of 100$ agricultural products is assigned to 15$ products for reproducing, 20$ products being sold to manufacturing, 30$ products to service, and the last 35$ products being used to meet external demand. Similarly, it can be seen from the second column that for the total output of 200$, the manufacturing needs to invest 20$ of agricultural products, 10$ of its own products, and 60$ of service input.

For convenience of analysis, coupled with the input-output table, the so-called technology input-output table, can be constructed to show the amount that each sector, for the purpose of producing one unit of its own product, needs for consuming the other sector's. The quantity in the technical input-output table represents the input coefficient t_{ij} (in Leontief's model) or consumption coefficient (in CPE model) of the economy, which can be obtained from Table 1, for example, the data of each column in Table 1 being divided by the total agriculture output 100 gives the agriculture sector's input coefficients and so on. Thus, the technology input table or consumption coefficient table is completed and shown in Table 2.

By Table 2, the input matrix T or consumption coefficient matrix A should be

$$T = A = \begin{bmatrix} 0.15 & 0.10 & 0.20 \\ 0.30 & 0.05 & 0.30 \\ 0.30 & 0.30 & 0.00 \end{bmatrix}. \tag{31}$$

The input-output table can comprehensively and systematically reflect the input-output relationship among all sectors of the national economy and reveal the economic and technological relations of interdependence and mutual restriction among all sectors in the production process. On the other hand, it can tell people about the output of various sectors of the national economy and how the output of these sectors is distributed to other sectors for production or to residents and society for final consumption or export abroad. Furthermore, it can tell people how each department obtains intermediate inputs and initial inputs from other departments for its own production.

Example 1 only provides the fundamental principle for the compilation of input-output table with a simple case. In general, the actual input-output table of an economy is much more complicated than this one provided and it generally includes national table, regional table, sectoral table, and joint enterprise table according to different scopes, as well as static table and dynamic table according to the model characteristics. In addition, there are input-output tables for

Table 1: Input-output table.

Input	Output				
	Agriculture	Manufacturing	Service	External demand	Total output
Agriculture	15 (x_{11})	20 (x_{12})	30 (x_{13})	35 (d_1)	100 (x_1)
Manufacturing	30 (x_{21})	10 (x_{22})	45 (x_{23})	115 (d_2)	200 (x_2)
Service	20 (x_{31})	60 (x_{32})	0 (x_{33})	70 (d_3)	150 (x_3)

Table 2: Technology input-output table.

Input	Output		
	Agriculture	Manufacturing	Service
Agriculture	0.15 (t_{11})	0.10 (t_{12})	0.20 (t_{13})
Manufacturing	0.30 (t_{21})	0.05 (t_{22})	0.30 (t_{23})
Service	0.20 (t_{31})	0.30 (t_{32})	0.00 (t_{33})

studying special issues such as environmental protection, population, and resources.

The function of a nation's input-output report is not only to reflect the direct and obvious economic and technological relations among various departments in the production process, but also to reveal the indirect, hidden, and even neglected economic and technological relations among various departments. The input-output table provides the basis for studying the industrial structure, especially for making and checking the national economic plan, studying the price decision, and conducting various quantitative analyses.

Example 2. Let A be a matrix like the following:

$$A = \begin{bmatrix} \frac{1}{10} & 0 & \frac{2}{5} \\ 0 & \frac{1}{5} & 0 \\ 0 & 0 & \frac{2}{5} \end{bmatrix}, \tag{32}$$

and it is easy to find the inverse of A as follows:

$$B = A^{-1} = \begin{bmatrix} 10 & 0 & -10 \\ 0 & 5 & 0 \\ 0 & 0 & \frac{5}{2} \end{bmatrix}. \tag{33}$$

A simple computation gives us the following eigenvectors X and the corresponding eigenvalues $\lambda_i\,(i = 1, 2, 3)$ of A:

$$X = \begin{bmatrix} 1 & 0 & \frac{4}{5} \\ 0 & 1 & 0 \\ 0 & 0 & \frac{3}{5} \end{bmatrix}, \tag{34}$$

$$\lambda = (\lambda_1, \lambda_2, \lambda_3) = \left(\frac{1}{10}, \frac{1}{5}, \frac{2}{5}\right).$$

It is easy to check that $1 - a_{ii} > \sum_{j \neq i} a_{ij}, i, j = 1, 2, 3$; this means the conditions of Theorem 1 are satisfied and then, by Theorem 1, CPE system (4) with consumption coefficient matrix A must have a growth balanced solution. One thing that needs special attention, here, is that the matrix A is reducible which is different from the fact that most of the existing dynamic properties for CPE and FFCME are based on the hypothesis of matrix A to be irreducible.

7. Conclusion

Other than the "planned economic" policy that had been put into practice in China for nearly 30 years from 1949 to 1978, due to the first industrial revolution which started at the sixties of the 18th century, there is the "fully free competition marketing economic" policy being conducted by the West for over one hundred years. There have been a tremendous number of research papers investigating various economic behaviors of the fully free competition marketing systems in the past hundred years but few studies investigated the planned economic systems till Professor Loo-Keng Hua, a world famous Chinese mathematician, presented a series of research papers in the 80s of the last century.

In parallel with the extensive investigations being done for various FFCME systems and motivated by studying along the direction of dynamical properties of a system, we have carried out some important dynamical properties for CPE system which were first proposed by the famous Chinese mathematician Loo-Keng Hua in this paper.

The main results obtained here include the followings: first, we obtain new conditions for the existence of growth balanced solution to the CPE system and investigate the stability of the balanced output as well; second, we propose a more generalized price equation and study the stability of the balanced price as well. All these results provide effective aspects for controlling the CPE system so as to keep the economy grow and develop stably.

Certainly, further study of the CPE system should be encouraged in two directions: the first one is that more suitable model is needed to match the economy developing state under the varying economy policy; for instance, each class's producing rate may be different but, here, in case of simplicity, with a hypothesis being the same growing rate; the second is that a deeper study, especially, the simulation of any economic model with the practical economy situation, will be more competitive than others. Just for this reason, we will aim at these problems in our future study on the economic systems.

Data Availability

No data were used for this paper.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The research was supported by Grant GD17XYJ29 from the Office of Philosophy and Social Science Research Project of Guang Dong Province, China.

References

[1] W. Leontief, "Dynamic Analysis in Studies in the Structure ofthe American Economy: Theoretical and Empirical ExplorationsInpinput-Output Analysis," in *Isard and Helen Kistin*, W. Leontief, H. B. Chenery, P. G. Clark et al., Eds., pp. 53–90, Oxford University Press, Oxford, UK, 1953..

[2] W. Leontief, "Lags and the stability of dynamic systems," *Econometrica*, vol. 29, no. 4, pp. 659–669, 1961.

[3] W. Leontief, "Lags and the stability of dynamic systems: a rejoinder," *Econometrica*, vol. 29, no. 4, pp. 674-675, 1961.

[4] W. Leontief, *Input-output Economics*, Oxford University Press, Oxford, UK, 1966.

[5] J. D. Sargan, "The instability of the Leontief dynamic model," *Econometrica*, vol. 26, no. 3, pp. 381–392, 1958.

[6] W. Leontief, "The dynamic inverse," in *Proceedings of the Contributions to input-output analysis: proceedings of the fourth international conference on input-output techniques*, North Holland, Geneva, Amsterdam, January 1970.

[7] A. Brody and A. P. Carter, "Inputcoutput techniques," in *Proceedings of the Fifth International Conference on inputCoutput Techniques*, North Holland, Geneva, Amsterdam, January 1972.

[8] A. P. Schinnar, "The Leontief dynamic generalized inverse," *The Quarterly Journal of Economics*, vol. 92, no. 4, pp. 641–652, 1978.

[9] J. Tsukui and Y. Murakami, *Turnpike Optimality in Input-Output Systems*, North Holland, Geneva, Amsterdam, 1979.

[10] R. S. Preston, "The wharton long-term model: input-output within the context of a macro forecasting model," *Econometric Model Performance: Comparative Simulations of the U.S. economy*, pp. 271–287, University of Pennsylvania Press, Philadelphia, PA, USA, 1976.

[11] W. Leontief, A. P. Carter, and P. A. Petri, *The Future of the World Economy: A United Nations Study*, Oxford University Press, Oxford, UK, 1977.

[12] V. Bulmer-Thomas, *InputCoutput Analysis in Developing Countries: Sources, Methods and Applications*, Wiley, Hoboken, NJ, USA, 1982.

[13] W. Leontief and F. Duchin, *The Future Impact of Automation on Workers*, Oxford University Press, Oxford, UK, 1985.

[14] W. Peterson, *Advances in Input-Output analysis:Technology, Planning, and Development*, Oxford University Press, Oxford, UK, 1991.

[15] R. E. Miller and P. D. Blair, *Input-output Analysis: Foundations and Extensions*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1985.

[16] L.-K. Hua, "On the mathematical theory of globally optimal planned economic systems," *Proceedings of the National Academy of Science, USA*, vol. 81, no. 20, pp. 6549–6553, 1984.

[17] L.-K. Hua, "Mathematical theory of large-scale optimization in planned economy (I)," *Chinese Science Bulletin*, vol. 29, no. 12, pp. 6–11, 1984, in Chinese.

[18] L.-K. Hua, "Mathematical theory of large-scale optimization in planned economy (II-III)," *Chinese Science Bulletin*, vol. 29, no. 13, pp. 769–772, 1984, in Chinese.

[19] L.-K. Hua, "Mathematical theory of large-scale optimization in planned economy (IV-VI)," *Chinese Science Bulletin*, vol. 29, no. 16, pp. 961–965, 1984, in Chinese.

[20] L.-K. Hua, "Mathematical theory of large-scale optimization in planned economy (VII)," *Chinese Science Bulletin*, vol. 29, no. 18, pp. 1089–1092, 1984, in Chinese.

[21] L.-K. Hua, "Mathematical theory of large-scale optimization in planned economy (VIII)," *Chinese Science Bulletin*, vol. 29, no. 21, pp. 1281-1282, 1984, in Chinese.

[22] L.-K. Hua, "Mathematical theory of large-scale optimization in planned economy (IX)," *Chinese Science Bulletin*, vol. 30, no. 1, pp. 1-2, 1985, in Chinese.

[23] L.-K. Hua, "Mathematical theory of large-scale optimization in planned economy (X)," *Chinese Science Bulletin*, vol. 30, no. 9, pp. 641–645, 1985, in Chinese.

[24] O. Blanchard, *The Need for Different Classes of Macroeconomic Models, Blog Post*, Peterson Institute for International Economics, Washington, D.C, USA, 2017, https://piie.com/blogs/.

[25] A. P. Schinnar, "A multidimensional accounting model for demographic and economic planning interactions," *Environment and Planning A: Economy and Space*, vol. 8, no. 4, pp. 455–475, 1976.

[26] G. Chinchilnisky, G. Heal, and A. Beltratti, "The green golden rule," *Economics Letters*, vol. 49, pp. 174–179, 1995.

[27] G. Chinchilnisky, "What is sustainable development?" *Land Economics*, vol. 73, pp. 476–491, 1997.

[28] C. Cinzia, J. Raja, and M. Simone, "Multi-criteria decision analysis with goal programming in engineering, management

and social sciences: a state-of-the art review," *Annals of Operations Research*, vol. 251, 2015.

[29] R. M. Solow and P. A. Samuelson, "Balanced growth under constant returns to scale," *Econometrica*, vol. 21, no. 3, pp. 412–424, 1953.

[30] A. H. Roger and R. Johnson, *Matrix Analysis*, Cambridge University Press, London, UK, 2nd edition, 2013.

Hindawi
Discrete Dynamics in Nature and Society
Volume 2020, Article ID 3198783, 9 pages
https://doi.org/10.1155/2020/3198783

Research Article

An Automatic System for Atrial Fibrillation by Using a CNN-LSTM Model

Fengying Ma,[1] **Jingyao Zhang**,[1] **Wei Chen**,[2,3] **Wei Liang**,[1] **and Wenjia Yang**[2,3]

[1]*School of Electrical Engineering and Automation, Qilu University of Technology (Shandong Academy of Sciences), Jinan, China*
[2]*School of Mechanical Electronic & Information Engineering, China University of Mining and Technology-Beijing, Beijing 100083, China*
[3]*School of Computer Science & Technology, China University of Mining and Technology, Xuzhou 221116, China*

Correspondence should be addressed to Wei Chen; chenwdavior@163.com and Wei Liang; dzhlw0918@qlu.edu.cn

Received 18 June 2020; Revised 24 July 2020; Accepted 27 July 2020; Published 28 August 2020

Guest Editor: Chi-Hua Chen

Atrial fibrillation (AF) is a common abnormal heart rhythm disease. Therefore, the development of an AF detection system is of great significance to detect critical illnesses. In this paper, we proposed an automatic recognition method named CNN-LSTM to automatically detect the AF heartbeats based on deep learning. The model combines convolutional neural networks (CNN) to extract local correlation features and uses long short-term memory networks (LSTM) to capture the front-to-back dependencies of electrocardiogram (ECG) sequence data. The CNN-LSTM is feeded by processed data to automatically detect AF signals. Our study uses the MIT-BIH Atrial Fibrillation Database to verify the validity of the model. We achieved a high classification accuracy for the heartbeat data of the test set, with an overall classification accuracy rate of 97.21%, sensitivity of 97.34%, and specificity of 97.08%. The experimental results show that our model can robustly detect the onset of AF through ECG signals and achieve stable classification performance, thereby providing a suitable candidate for the automatic classification of AF.

1. Introduction

Heart disease is the leading cause of human death, and the number of deaths due to cardiovascular diseases accounts for a large proportion of the total number of deaths worldwide [1]. Most cardiovascular diseases are often accompanied by arrhythmia. Among them, atrial fibrillation (AF) is the most common persistent arrhythmia. In our country, there are more than 10 million people suffering from AF. Its incidence increases with age, but in recent years, its incidence has shown an increasing trend in people of younger age groups [2–4]. Simultaneously, a series of complications related to AF, such as stroke, heart failure, and other diseases, also lead to high morbidity and mortality [5–7]. However, AF also shows strong unpredictability, and capturing AF signals in real time is difficult [8, 9]. Electrocardiogram (ECG) detection technology forms an important basis for AF diagnosis [10–12]. Therefore, the application of automatic detection technology for diagnosing AF is necessary. Consequently, machine learning has significantly contributed to the development of real-time monitoring of AF, and timely intervention in the effective detection of AF can avoid serious consequences caused by an exacerbation of the disease [13].

With developments in information technology and artificial intelligence technology, the automatic classification of AF has made great progress. The traditional ECG classification algorithm is composed of two parts: feature extraction and classifier. First, principal component analysis, latent Dirichlet allocation, and other methods are designed for feature extraction and then placed into a support vector machine or random forest, among others, for classification in the classifier. A complete ECG signal is shown in Figure 1 where AF is mainly seen by the disappearance of the p wave or irregular RR intervals. The RR interval is the amount of time change between two R waves, and the feature recognition method based on the RR interval shows high accuracy. Because the wave value of the R wave has the largest

peak in the ECG signal, locating it is easy. However, the low amplitudes of the P and T waves make it challenging for them to be detected, and the feature extraction algorithm is still not mature enough. With developments in machine learning, traditional models have some insurmountable defects. First, traditional algorithms need to design feature extraction methods to extract useful information and combine machine learning algorithms for classification. This process may be accompanied with the loss of some information. When the extracted features cannot fully reflect the data, the classification results may appear to have larger errors. Second, it mainly relies on considerable prior expert knowledge and sufficient biomedical signal processing capabilities. On this basis, designing a good classifier algorithm is also necessary, but achieving optimal results is difficult.

Unlike traditional machine learning algorithms, deep learning-based methods have the ability to mine complex relations, and useful features of data and have been widely researched and applied in the automatic classification of AF. Xu et al. proposed a framework that combined an improved frequency slice wavelet transform and a convolutional neural network (CNN) for automatic AF beat recognition and achieved good performance [14]. Wei et al. constructed a synchronous feature of each heartbeat of an ECG signal through a recursive complex network and subsequently used CNN to detect AF by analyzing the eigenvalues of the recursive complex network [15]. Andersen et al. proposed an end-to-end model combining CNN and recurrent neural networks to classify ECG signals as AF or a normal sinus rhythm [11]. Pourbabaee et al. developed a deep learning machine to screen and identify patients with paroxysmal AF [16]. Dang et al. proposed a model that uses a CNN-BLSTM network to diagnose arrhythmia and uses ECG signals to automatically detect AF, achieving relatively good results [17]. Several researchers have shown that combining the deep learning features with the classifier will significantly improve the performance of the system and make the classification results more ideal. Although the above research can effectively solve the classification problem of AF, we can observe that various neural networks have the ability to extract complex nonlinear characteristics from the original data without human intervention; however, learning the thinking mechanism of the ECG signal features with high accuracy required for monitoring is still a difficult task. CNN and long short-term memory networks (LSTM) are very efficient for feature extraction of ECG signals, and this superiority is applied herein to the AF detection algorithm.

In this study, we proposed a new diagnosis method for AF named CNN-LSTM, which can automatically detect AF from ECG signals. The contributions of this study are as follows:

(i) We propose an automatic recognition method named CNN-LSTM that uses heartbeat features as input datasets to automatically identify AF in an ECG signal.

(ii) CNN has advantages in image processing, while LSTM can compensate for the shortcomings of CNN in the context sequence. Therefore, the combination of CNN and LSTM can effectively improve accuracy in the field of AF recognition.

(iii) The use of multiscale signals representing the AF characteristics as the input of the network reduces computing resources. The design can be used to extract multiscale features and improve the generalization ability of the network model, and this study provides a high-precision classification method to meet the real-time monitoring needs of AF.

In the following sections, we provide a detailed experimental process and verify the performance of the method in the open access database. This automated method can analyze a large amount of data in a short time while ensuring high accuracy; thus, it may become a practical tool for providing real-time monitoring for patients and reducing the work pressure on doctors.

Section 1 of this article deals with the background of current research on AF and related research algorithms that have been implemented. Section 2 covers the data source and related network structure required for the experiment. Section 3 presents the experimental details, results, and analysis of the results. Section 4 presents the concluding statements and prospects for future work.

2. Material and Methods

2.1. Description of Dataset. Our experimental research is based on data from the MIT-BIH Atrial Fibrillation Database, which is publicly available from PhysioNet [18, 19]. This database includes 25 long-term ECG Holter records from different subjects (mainly paroxysmal attacks). It contains two ECG signal channels with AF annotations. The sampling rate of this database is 250 Hz, and these records also include beat notes manually marked by expert clinicians.

We preprocess the ECG signal to train and evaluate the automatic AF prediction method based on the CNN-LSTM model. As the duration of the AF recording in these data is different from the normal recording duration, all ECG signals are divided into the same duration and data balance is performed to better apply the learning of the model. After segmentation, 960,000 short-term ECG segments were obtained, comprising 480,000 segments of AF records and 480,000 segments of normal records. Figure 2 shows a comparison of the AF signal and the normal signal. The ECG signal segment is divided into a training set and a test set. To better detect the classification effect of the model, the signal processing and segmentation of the dataset are random in nature.

2.2. Networks. This section first reviews the CNN and LSTM network models, which are closely related to the model structure proposed herein. Then, our proposed research model is put forward, and the structure, parameters, and mathematical expressions of the model are described in detail.

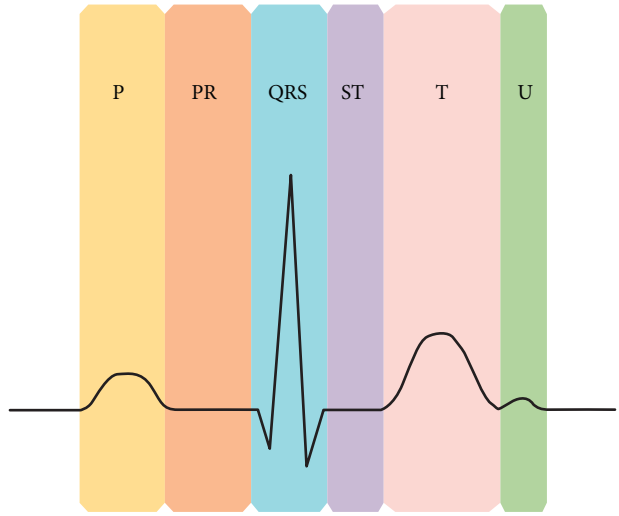


Figure 1: A complete ECG signal.

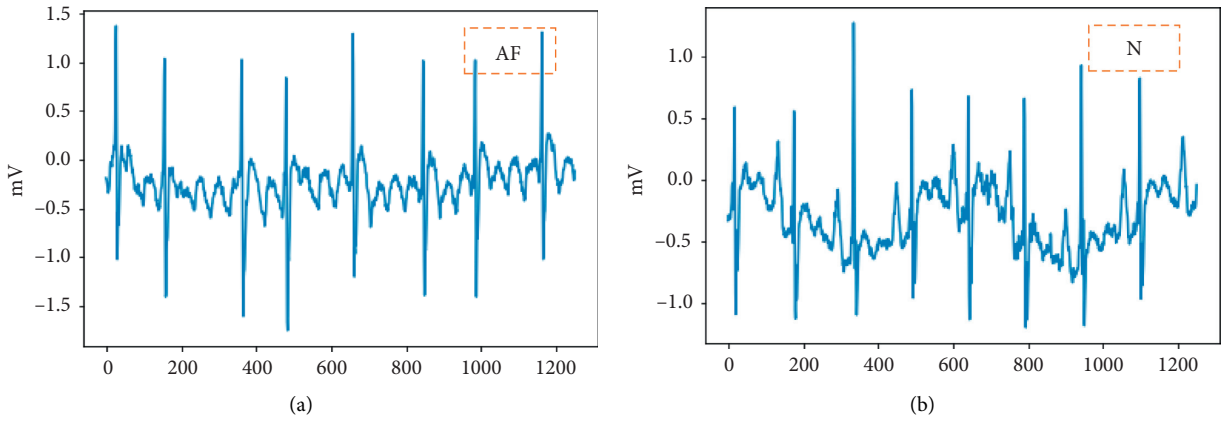


Figure 2: Comparison of AF signal and N signal.

In this section, we describe the network structure model proposed in this paper, which mainly includes two convolutional layers, one LSTM layers, fully connected layers, and other computing operations.

2.2.1. Convolutional Neural Network. CNN is a feedforward neural network, and it mainly includes an input layer, a convolutional layer, a pooling layer, and an output layer. Its special network structure has great advantages in feature extraction and learning, especially in the field of image recognition, and thus it can achieve great success. Its structure is shown in Figure 3.

The CNN is connected to the input layer through a convolution kernel. The convolution kernel performs dot multiplication through a sliding window to achieve multi-scale feature extraction. Simultaneously, the weight-sharing mechanism of the convolution layer makes it more effective for feature extraction, greatly reducing the number of free variables that need to be learned. Subsequently, we add a pooling layer after the convolutional layer to reduce the feature matrix and network complexity. Because the input ECG signals are one-dimensional time series, we use one-dimensional convolution in the convolution layer, as shown in Figure 4.

Before the data training, we normalized the data. The convolutional layer extracts features from the original input. The output of the a-th neuron of the one-dimensional convolutional layer is shown in the following equation:

$$O_a = \delta\left(\sum_{j=1}^{n} W_j X_{a-j+n} + b\right). \tag{1}$$

The input sequence is $X_l (l = 1, 2, \ldots, n)$, where W denotes a matrix of weight coefficients, b is an offset coefficient, and n is the number of convolution kernels. Then, the result of the convolution is input into an activation function δ (in this case ReLU), and then the result of the convolution layer is fed back to the pooling layer.

2.2.2. Long Short-Term Memory Network. LSTM is a special recurrent neural network. LSTM is suitable for applications such as natural language processing [20, 21] and biomedical

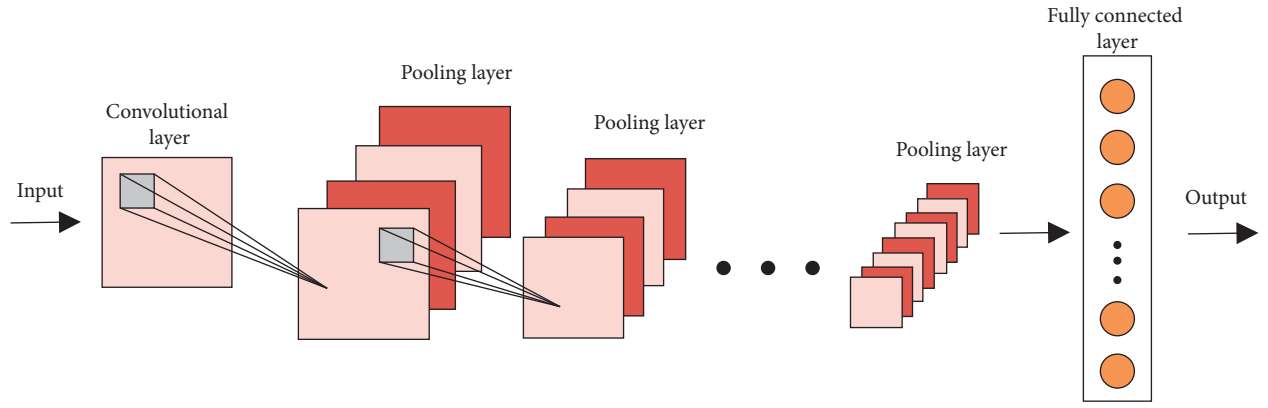


Figure 3: CNN structure.

signal processing [22, 23]. LSTM improves the standard RNN model and adds a gate mechanism. It overcomes the problems of gradient disappearance, gradient explosion, and length dependence of traditional RNNs. The hidden layer of LSTM comprises an input gate, a forget gate, and an output gate. The structure is shown in Figure 5.

The input of the LSTM hidden layer includes not only the input X_t of the current sequence but also the state C_{t-1} of the hidden layer at the previous time; then, the output vector h_{t-1} and the output h_t and C_t of the current state are calculated. The key part of LSTM is what information we will discard from the cell state C_{t-1} at the last moment and how much information can be transferred to the current state C_t. This decision is made through a forget door. The next step is to decide how much new information is added to the next state. Finally, we determine the output value based on the current C_t and h_t. The update method of LSTM is as follows:

$$\begin{aligned} f_t &= \delta\left(W_f\left[h_{t-1}, X_t\right] + b_f\right), \\ i_t &= \delta\left(W_i\left[h_{t-1}, X_t\right] + b_i\right), \\ j_t &= \tanh\left(W_c\left[h_{t-1}, X_t\right] + b_c\right), \\ O_t &= \delta\left(W_o\left[h_{t-1}, X_t\right] + b_o\right), \\ h_t &= O_t * \tanh\left(C_t\right), \end{aligned} \tag{2}$$

where C_t is the state information of the memory unit, j_t is the accumulated information at the current moment, W is the weight coefficient matrix, b is the bias term, σ is the sigmoid activation function, and tanh is the hyperbolic tangent activation function.

2.2.3. Proposed Architecture. Neural networks have their own unique feature learning method. The CNN model converts the original input into a fixed-length vector representation through convolution kernels, sliding windows, and pooling to capture local features in the input, but the original data arrive. The dependency relation is difficult to learn, and LSTM can better understand the content of the input information through the memory unit that can compensate for the defects of the CNN. Therefore, a CNN-LSTM deep learning model is proposed to achieve the automatic classification of AF in this paper [24, 25].

Figure 6 shows the proposed CNN-LSTM network architecture. After inputting the ECG signal, the convolutional layer and the pooling layer in the CNN first extract local features and subsequently enter the hidden layer of LSTM to obtain optimal feature representation. Finally, the nonlinear function softmax in the fully connected layer is classified into the corresponding categories. CNN adds some processing such as normalization that can avoid overfitting and speed up training.

3. Experimental Results

3.1. Performance Evaluation. To estimate the performance of heartbeat classification, the performance of the model is usually evaluated with accuracy, specificity, and sensitivity [26–28]. They are defined as follows:

$$\begin{aligned} \text{sensitivity} &= \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\%, \\ \text{specificity} &= \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100\%, \\ \text{accuracy} &= \frac{\text{TN} + \text{TP}}{\text{TN} + \text{FP} + \text{FN} + \text{TP}} \times 100\%, \end{aligned} \tag{3}$$

where TP denotes the number of correctly classified AF signals; FP denotes the number of incorrectly classified AF signals; TN is the number of correctly classified N signals; and FN is the number of misclassified N signals. The results of the experimental study are presented in the next section.

3.2. Implementation Details and Results. The experiment herein uses the TensorFlow neural network framework. Before the experiment started, the data labels were converted into corresponding one-hot vectors. The experiment is based on an equal number of AF records and normal records for training and testing, and the signal processing of the dataset is random. During the experiment, parameter optimization was performed, the batch size was set to 128, the learning rate was 0.01, multiple iteration training was performed, and the Adam updater was used to update the weights to obtain the best classification results. Table 1 lists the relevant parameters of the experimental network.

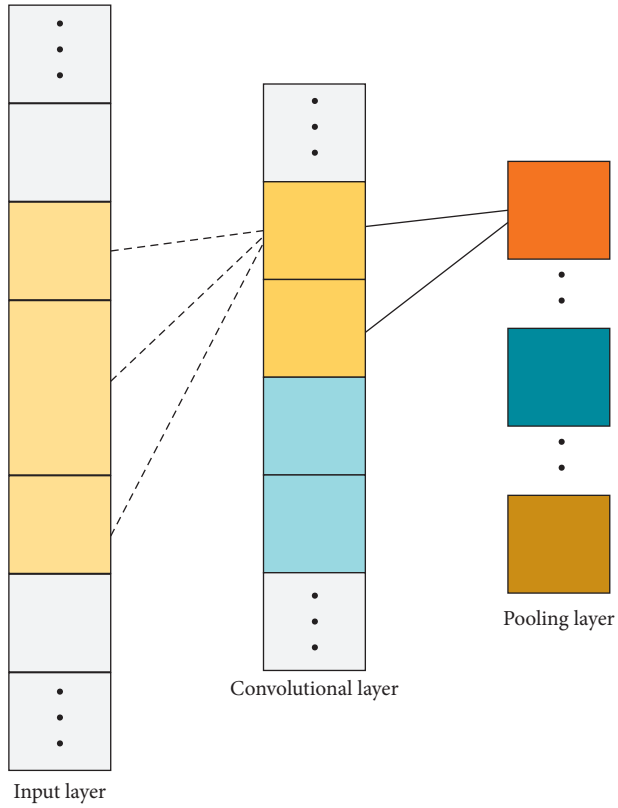


Figure 4: One-dimensional convolutional neural network structure.

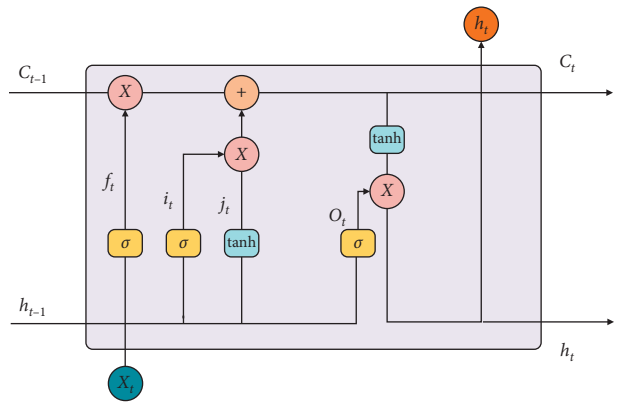


Figure 5: LSTM hidden layer structure diagram.

Figure 7 shows the loss and accuracy curves of the CNN-LSTM model. It indicates performance change in the training set as the number of iterations increases. The network continues to converge, and the model does not appear to be overfitting.

Figure 8 shows the receiver operating characteristic curve (ROC) curve of the test set. The abscissa of the curve is the false positive rate, and the ordinate is the true positive rate. AUC denotes the area under the ROC curve. The AUC realized by the model was 0.97. The closer the AUC value is to 1, the better is the performance of the model.

Figure 9 shows the confusion matrix of the test set, which is used to measure the accuracy of a classifier. The upper left corner is TP, the upper right corner is FP, the lower left corner is FN, and the lower right corner is FP. We can convert the result of the quantity in the confusion matrix to a ratio between 0 and 1 to facilitate standardized measurements.

As mentioned above, the CNN-LSTM model achieved an overall classification accuracy of 97.28% on the test set, with a sensitivity of 97.51% and a specificity of 97.06%.

In this study, by combining the deep learning model of CNN and LSTM, i.e., using CNN and LSTM to extract the characteristics of the ECG, the model can automatically extract features and achieve higher accuracy.

Table 2 shows a series of scientific studies based on ECG signals in the MIT-BIH AF database. It mainly includes three evaluation indices: accuracy, sensitivity, and specificity.

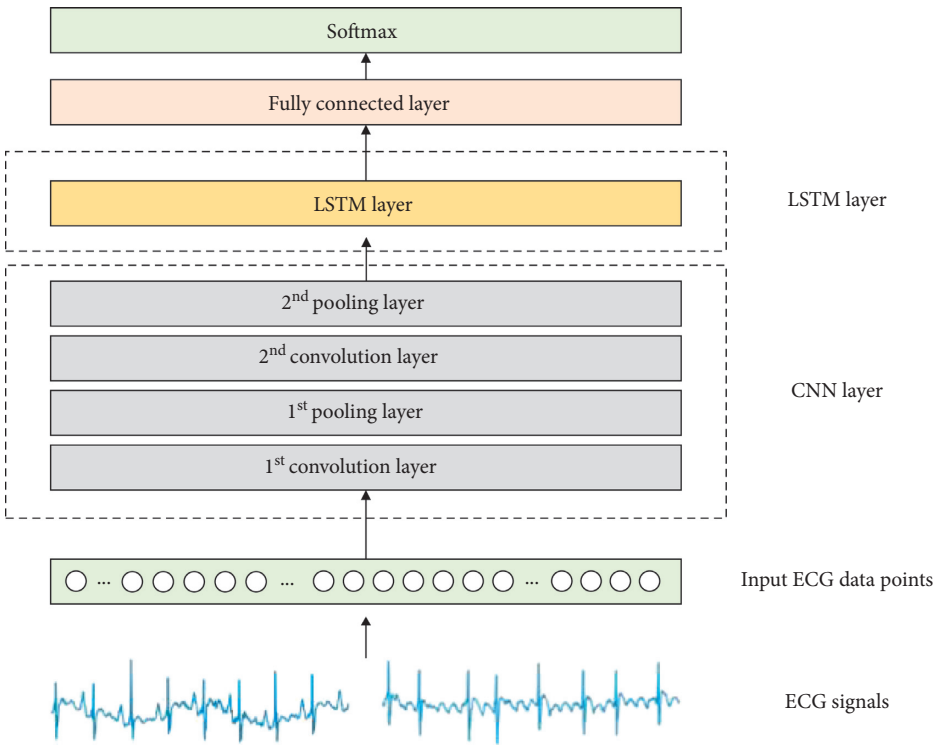


Figure 6: CNN-LSTM network structure.

Table 1: The parameters of the model.

Network layer type	Kernel size	Strides
Input layer	—	—
Con1d	17	1
Pooling1d	5	5
Con2d	7	1
Pooling2d	5	5
LSTM	—	—
Fully connected layer	—	—
Softmax	—	—

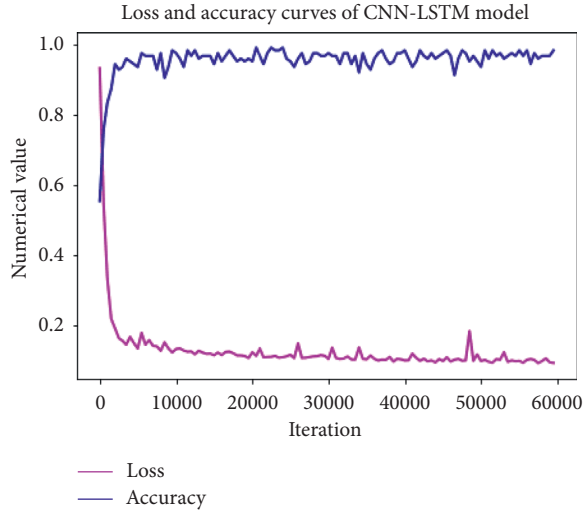


Figure 7: Loss and accuracy curves.

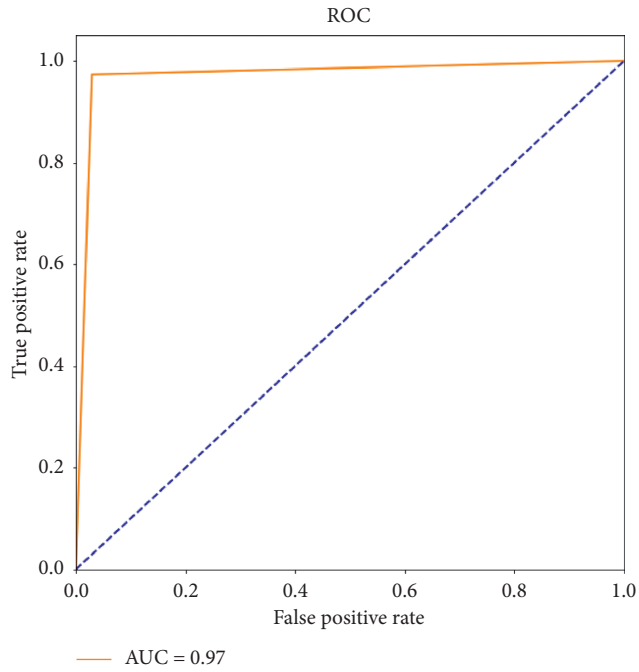


Figure 8: ROC curves.

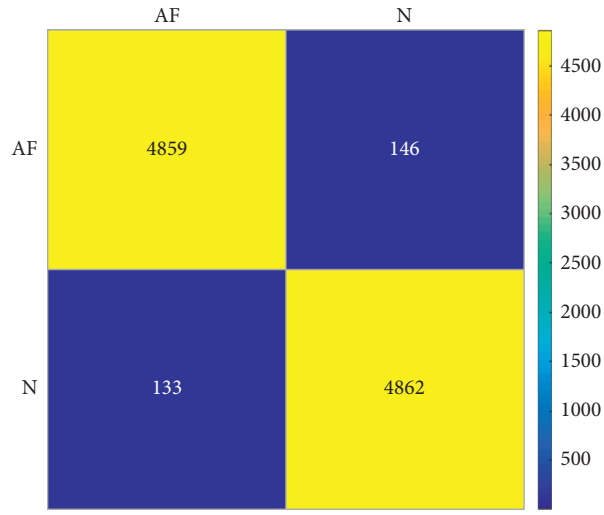


Figure 9: Confusion matrix of test set.

Table 2: Comparison with previous work on the MIT-BIH AF database.

Author	Database	Features	Classifier	Accuracy (%)	Sensitivity (%)	Specificity (%)
Xu et.al. [15]	AFDB	MESWT	CNN	85.85	79.05	89.99
Wei et.al. [14]	AFDB	RCN	CNN	94.59	94.28	94.91
Andersen et.al. [11]	AFDB MITDB NSRDB	RRI	LSTM + CNN	87.4	98.6	86.4
Dang et.al. [17]	AFDB	RR	CNN-BLSTM	96.59	99.93	97.03
		P-QRS-T	CNN-LSTM	94.07	94.25	92.73
Proposed model	AFDB	Deep features	CNN-LSTM	97.21	97.34	97.08

Through a comparison, we can observe that out proposed CNN-LSTM network model has improved on the input signal of the model and network structure compared with other deep learning methods and achieved good results.

4. Conclusion

In this study, we conducted an in-depth study of the ECG classification algorithm and constructed a network

combining CNN and LSTM. This network can extract the characteristics of ECG signals and classify them. Compared with traditional ECG classification methods, our proposed CNN-LSTM network structure used the MIT-BIH AF database and achieved a high classification accuracy. The experimental results confirm that our proposed CNN-LSTM network is effective for the automatic detection and classification of AF. In addition, this method occupies fewer computing resources and can theoretically achieve real-time performance, thereby contributing to the development of wearable ECG detection devices. Our future research may involve the use of a model that classifies AF tasks under nonfixed scale inputs to achieve further optimization of the neural network.

Data Availability

The data used to support the findings of this study have not been made available because the data also form part of an ongoing study. The original data of the study can be obtained at https://physionet.org/.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported by the Shandong University Undergraduate Teaching Reform Research Project (approval number: M2018X078) and the Shandong Province Graduate Education Quality Improvement Program 2018 (approval number: SDYAL18088). This study was also partially supported by the Major Science and Technology Innovation Projects of Shandong Province (grant no. 2019JZZY010731).

References

[1] H. Wang, N. Mohsen, A. Christine et al., "Global, regional, and national life expectancy, all-cause mortality, and cause-specifc mortality for 249 causes of death, 1980–2015: a Systematic Analysis for the Global Burden of Disease Study 2015," *The Lancet*, vol. 388, no. 10053, pp. 1459–1544, 2016.

[2] C. R. C. Wyndham, "Atrial fibrillation: the most common arrhythmia," *Texas Heart Institute Journal*, vol. 27, no. 3, pp. 257–267, 2000.

[3] C.-H. Chen, F. Song, F.-J. Hwang, and L. Wu, "A probability density function generator based on neural networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 541, Article ID 123344, (2020).

[4] V. Markides and R. J. Schilling, "Atrial fibrillation: classification, pathophysiology, mechanisms and drug treatment," *Heart*, vol. 89, no. 8, pp. 939–943, 2003.

[5] L. Mainardi, Sornmo, and S. Cerutti, *Understanding Atrial Fibrillation: The Signal Processing Contribution*, Morgan Claypool Publishers, San Rafael, CA, USA, 2008.

[6] E. J. Benjamin, P. A. Wolf, R. B. D'Agostino, H. Silbershatz, W. B. Kannel, and D. Levy, "Impact of atrial fibrillation on the risk of death," *Circulation*, vol. 98, no. 10, pp. 946–952, 1998.

[7] C.-H. Chen, "A cell probe-based method for vehicle speed estimation," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E103.A, no. 1, pp. 265–267, 2020.

[8] C.-H. Chen, F.-J. Hwang, and H.-Y. Kung, "Travel time prediction system based on data clustering for waste collection vehicles," *IEICE Transactions on Information and Systems*, vol. E102.D, no. 7, pp. 1374–1383, 2019.

[9] J. R. Mehall, R. M. Kohut Jr., E. W. Schneeberger, W. H. Merrill, and R. K. Wolf, "Absence of correlation between symptoms and rhythm in "Symptomatic" atrial fibrillation," *The Annals of Thoracic Surgery*, vol. 83, no. 6, pp. 2118–2121, 2007.

[10] S. George, I. Rodriguez, D. Ipe, P. T. Sager, I. Gussak, and B. Vajdic, "Computerized extraction of electrocardiograms from continuous 12- lead holter recordings reduces measurement variability in a thorough QT study," *The Journal of Clinical Pharmacology*, vol. 52, no. 12, pp. 1891–1900, 2012.

[11] R. S. Andersen, A. Peimankar, and S. Puthusserypady, "A deep learning approach for real-time detection of atrial fibrillation," *Expert Systems with Applications*, vol. 115, pp. 465–473, 2019.

[12] S. Asgari, A. Mehrnia, and M. Moussavi, "Automatic detection of atrial fibrillation using stationary wavelet transform and support vector machine," *Computers in Biology and Medicine*, vol. 60, pp. 132–142, 2015.

[13] Z. Yao, Z. Zhu, and Y. Chen, "Atrial fibrillation detection by multi-scale convolutional neural networks," in *Proceedings of the 2017 International Conference on Information Fusion IEEE*, Xi'an, China, July 2017.

[14] X. Xu, S. Wei, C. Ma, K. Luo, L. Zhang, and C. Liu, "Atrial fibrillation beat identification using the combination of modified frequency slice wavelet transform and convolutional neural networks," *Journal of Healthcare Engineering*, vol. 2018, Article ID 2102918, 8 pages, 2018.

[15] X. J. Wei, C. Zhang, M. Liu et al., "Atrial fibrillation detection by the combination of recurrence complex network and convolution neural network," *Journal of Probability and Statistics*, vol. 2019, Article ID 8057820, 9 pages, 2019.

[16] B. Pourbabaee, M. J. Roshtkhari, and K. Khorasani, "Deep convolutional neural networks and learning ECG features for screening paroxysmal atrial fibrillation patients," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 12, pp. 2095–2104, 2018.

[17] H. Dang, "A novel deep arrhythmia-diagnosis network for atrial fibrillation classification using electrocardiogram signals," *IEEE Access*, vol. 7, pp. 2169–3536, 2019.

[18] A. L. Goldberger, L. A. N. Amaral, L. Glass et al., "PhysioBank, PhysioToolkit, and PhysioNet," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[19] G. Moody, "A new method for detecting atrial fibrillation using rr intervals," *Computers in Cardiology*, vol. 10, pp. 227–230, 1983.

[20] W. De Mulder, S. Bethard, and M.-F. Moens, "A survey on the application of recurrent neural networks to statistical language modeling," *Computer Speech & Language*, vol. 30, no. 1, pp. 61–98, 2015.

[21] H. Palangi, L. Deng, Y. Shen et al., "Deep sentence embedding using long short-term memory networks: analysis and application to information retrieval," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 4, pp. 694–707, 2016.

[22] O. Faust, A. Shenfield, M. Kareem, T. R. San, H. Fujita, and U. R. Acharya, "Automated detection of atrial fibrillation using long short-term memory network with RR interval

signals," *Computers in Biology and Medicine*, vol. 102, pp. 327–335, 2018.

[23] P. Cao, X. Li, K. Mao et al., "A novel data augmentation method to enhance deep neural networks for detection of atrial fibrillation," *Biomedical Signal Processing and Control*, vol. 56, Article ID 101675, 2020.

[24] J. Lu, L. A. Hendricks, M. Rohrbach et al., "Long-term recurrent convolutional networks for visual recognition and description," *IEEE Transactions on Pattern Analysis Machine Intelligence*, vol. 39, no. 4, pp. 677–691, 2017.

[25] W.-P. XiongT.-C. Li et al., "Research on partial least squares method based on deep confidence network in traditional chinese medicine," *Discrete Dynamics in Nature and Society*, vol. 2020, Article ID 4142824, 10 pages, 2020.

[26] F. Ma, J. Zhang, W. Liang et al., "Automated classification of atrial fibrillation using artificial neural network for wearable devices," *Mathematical Problems in Engineering*, vol. 2020, Article ID 9159158, 6 pages, 2020.

[27] C. ZhangJ. He et al., "A crash severity prediction method based on improved neural network and factor Analysis," *Discrete Dynamics in Nature and Society*, vol. 2020, Article ID 4013185, 13 pages, 2020.

[28] D. M. W. Powers, "Evaluation: from precision, recall and ffactor to roc, informedness, markedness correlation," *Journal of Machine Learning Technologies*, vol. 2, pp. 37–63, 2011.

Hindawi
Discrete Dynamics in Nature and Society
Volume 2020, Article ID 2162869, 18 pages
https://doi.org/10.1155/2020/2162869

Research Article

Colored Petri Net-Based Verification and Improvement of Time-Sensitive Single-Unit Manufacturing for the Soil Preparation Instrument of Space Missions

Kai Leung Yung,[1] Ming Gao ,[2,3] An Liu,[3] Wai Hung Ip,[1,4] and Shancheng Jiang[5]

[1]*Department of Industrial and System Engineering, The Hong Kong Polytechnic University, Hong Kong SAR, China*
[2]*Center for Post-doctoral Studies of Computer Science, Northeastern University, Shenyang, China*
[3]*School of Management Science and Engineering, Dongbei University of Finance and Economics, Dalian, China*
[4]*University of Saskatchewan, College of Engineering, Saskatoon, Canada*
[5]*School of Intelligent Systems Engineering, Sun Yat-sen University, Guangzhou, China*

Correspondence should be addressed to Ming Gao; gm@dufe.edu.cn

Received 5 June 2020; Revised 17 July 2020; Accepted 20 July 2020; Published 25 August 2020

Guest Editor: Chi-Hua Chen

Various space missions, including the Russian and Chinese interplanetary exploration collaboration in 2011 and the Phobos-Grunt space project to be relaunched by the Chinese in 2025, carry a soil preparation system (SOPSYS), which is an instrument used for scientific experiments. The design and manufacture of this precision instrument require stringent manufacturing processes and workflow of the highest quality, with every process in the project carefully monitored and controlled. All processes should be completed within the deadline so that the space project can be launched at the scheduled time. The colored Petri net (CPN) modeling method can describe a variety of resource types and execution logic, and it can be formally verified. Based on these advantages, we clearly describe the complex structure of the SPOSYS unit production process. In addition, we use critical time and the 6 sigma system to evaluate the availability and reliability of workflows, and we use elimination and simplification (ECRS) methods and constraint theory to improve the manufacturing process of the SOPSYS unit. We further provide optimization theories, methods, and insights for workflow management in time-sensitive and independent manufacturing systems.

1. Introduction

Workflow is a case-based term focusing on the logistics part of a business. The purpose of studying it is to ensure that tasks are completed by the right people at the right time [1]. Such workflow has long been a trending topic since the introduction of information systems in the 1990s. At that time, information systems served as the tool to monitor, control, and support the business process [2]. Today, workflow-related research has been performed in a large number of industries; for example, airport traffic simulations [3], service industry [4], distributed manufacturing processes [5], supply chain networks [6], and plastic plants [7]. Although the topic of workflow is discussed in almost every industry, it has seldom been mentioned in space instrument manufacturing processes, which are extremely complicated and time-sensitive, and must be carefully managed. Hence, a study on workflow simulation analysis for space instrument manufacturing processes is imperative.

Petri net (PN) is a simulation tool developed by Dr. Carl Adam Petri in his doctoral paper. The classic Petri net, which often refers to the original one, has three elements: place, transition, and functions [8]. Mathematically, a Petri net can be represented with a tuple (P, T, and F). P stands for a finite set of places, T stands for a finite set of transitions, and F stands for a finite set of functions. Places and transitions map to buffers and activities. Functions involve the logical relationship between places and transitions, and tokens that

flow in the net can represent the state of the system. When dealing with a complex system, the classical Petri net is often insufficient for modeling such a system. The colored Petri net (CPN) introduced a new element—color—to the classical PN [9]. The attributes of token can now be specified through color, which helps exponentially enlarge the complexity of the system that CPN can model. As a fully developed modeling tool, CPN has been used in many industries, but its application in the space instrument manufacturing process is limited.

Phobos-Grunt is a Russian space mission that launched in 2011. However, the mission was not successful. One of its systems is the soil preparation system, which was developed by Hong Kong Polytechnic University [10]. The space mission is planned to refly in 2025, and it is, therefore, necessary to examine every process of the mission and ensure that the failure will not reoccur. This case, therefore, provides an opportunity to conduct research on the spacecraft manufacturing process workflow management using CPN.

This paper focuses on analyzing the workflow of a space instrument manufacturing process prior to the implementation stage through CPN modeling. It is imperative to build an appropriate workflow model at the abstract level in order to support process design and analysis [11]. Thus, the first focus of this paper is to present an appropriate CPN model for a specific space instrument manufacturing process. This model should be appropriate, comprehensive (considering both schedule and resources), and readable. Secondly, for the purpose of presenting a workflow improvement method, this paper will use the data derived from the case colored Petri net simulation to propose a new workflow using an existing project scheduling model. Then, another CPN model will be constructed to validate the performance of the proposal. The aim is to complete in time and improve the resource utilization.

This paper will contribute with respect to both practical and academic contexts. The practical contribution of this paper would be the optimization of the Phobos-Grunt SOPSYS delivery unit manufacturing process. The Phobos-Grunt space project is scheduled to relaunch in 2025. There are less than five years left to prepare for the relaunch; hence, every process in the project should be carefully monitored and controlled. All processes must be completed within the deadline so that the space project can be relaunched at the scheduled time. The academic contribution of this paper is that it has reference value for other time-sensitive unit manufacturing process optimization methods. The CPN modeling method we use can represent more complex manufacturing processes. For example, the CPN model used in this paper can clearly represent the delay time and rework structure. In addition, CPN's existing simulation software is more integrated, and the model can be formally verified [12]. This is something that traditional DAG and AOA cannot do.

2. Related Work

2.1. Workflow Optimization. Regarding the research of the workflow optimization method, Dewan et al. [13] focus on the task redesign process in workflow optimization. The study uses the case of IBM business information flow to demonstrate the technique. Dewan et al. argue that job bundling could be a useful approach to improve communication between tasks. This point is quite useful in this paper because during manufacturing processes, the rework delay can be effectively reduced through good communication. This is an example for task redesign. Another study focuses on the workflow optimization approach itself and presents a way to use FlowOpt in workflow optimization processes [14]. This process includes modeling, visualizing, analyzing, and optimizing the production workflow. Similarly, our paper will follow in the case modeling and analysis using software optimization of the workflow. Crop et al. also conducted a study on workflow optimization in the radiotherapy industry [15].

Studies in workflow optimization have been conducted in different industries. Dewan [13] optimized the workflow of business information flow. Karlik et al. [16] used CPN modeling and simulation methods to study the time and throughput of data-aware workflows to reflect the actual workload and gross profit of an enterprise over a period of time. Yanhua Du et al. [17] used Petri nets and neural networks in the development of an intelligent logic controller for an experimental manufacturing plant to provide the flexibility and intelligence required for this type of dynamic system. Kasemset [18] optimized the manufacturing process in a paper package factory. However, one industry has rarely been mentioned in the study of workflow optimization, the aerospace industry. Essentially, there is no comprehensive workflow optimization study that has been conducted in the aerospace industry.

Various qualitative and quantitative methods have been used to study workflow improvement. One paper package manufacturing process study used the ARENA program to model the workflow and evaluated the performance [18]. This is a sound approach in verifying the credibility of the optimization proposal, but the model built is too simple to reflect a real-world case. The study conducted by Dewan et al. [13] uses a quantitative validation method. It developed its own algorithm to calculate the delay per unit and cost per delay time unit. The algorithm is correct to some degree, but it still works in a relatively abstract manner that is rather difficult to understand. Crop et al. used Python-based discrete event simulation software to simulate the process and validated its ConWIP-based proposal [15]. The basic concept of this software is to logically model the process from a resource and time perspective. It is a reliable simulation tool but is not suitable for modeling concurrent systems.

In a study conducted by Lv et al. [5], the authors use a hierarchical CPN to model a distributed manufacturing process. In this study, Petri net's properties make it possible to simulate the concurrency and synchronization of a system. Hence, the hierarchical modeling approach can be achieved through a hierarchical CPN. This modeling technique is implemented in this paper.

Another common example of Petri net is in supply chains. Wang et al. presented a Petri net implementation method in a proactive resilient holistic supply chain network

[6]. It presents the specific CPN structure built for the supply chain network. This is a solid approach to show the establishing process of a CPN model. Although the study provides a highly detailed design process of the modeling, it is not validated through specific case data.

A CPN implication in a plastic manufacturing process workflow management was conducted by Fung et al. [7]. Their study placed substantial effort into describing how to map the workflow elements into the CPN model. This is also the main focus of this study. Although this study presents a highly detailed mapping theory regarding how to use CPN elements to represent workflow elements, there are no data to demonstrate that the model can actually generate real data, such as delay time or resource utilization.

A CPN simulation has been implemented in an airport pavement traffic case [3]. That study used the subnet concept to establish the CPN model for the case. It first introduced the concept of CPN and other extended CPNs, and then the model for the case was developed. The simulation data were then collected; although the simulation data showed a certain degree of flight delay, the study did not continue to optimize the workflow.

We have reviewed the workflow optimization literature related to this paper. A gap in the research area has been identified: there is no workflow optimization study for aerospace industry cases. Second, mixed qualitative and quantitative workflow optimization methods have seldom been demonstrated through simulations or algorithms. Third, we have summarized some effective viewpoints about the realization of CPNs in workflow optimization research, and we determined that discrete event simulation software that can simulate concurrent systems should be the best choice for modeling and analysis.

In this paper, we first introduced the basic concept of the CPN and extended CPNS. Then, we explained the mapping theory to map case workflow elements into CPN elements: the CPN model will be developed with a bottom-up approach that involves constructing the entire process from the substructure. Finally, we will collect and analyze several key data in the manufacturing process and propose improvement measures. The contribution of this study is that we make use of the relatively integrated software support of CPN, the color set of the theoretical model, time modeling and automatic verification, and other advantages to propose and verify improvement suggestions through case studies. Furthermore, the logic of this process is similar to the manufacturing process of most time-sensitive single units, and the same process and technology can be used in other manufacturing systems to achieve workflow optimization.

3. Methodology

The focus of this article is to use the CPN Tools simulation optimization method to optimize the manufacturing process of the SOPSYS delivery unit and provide a reference method and framework for such process optimization relating to space missions. The method and framework are shown in Figure 1. First, we sorted out the parameters involved in the problem and used CPN Tools to build an initial model

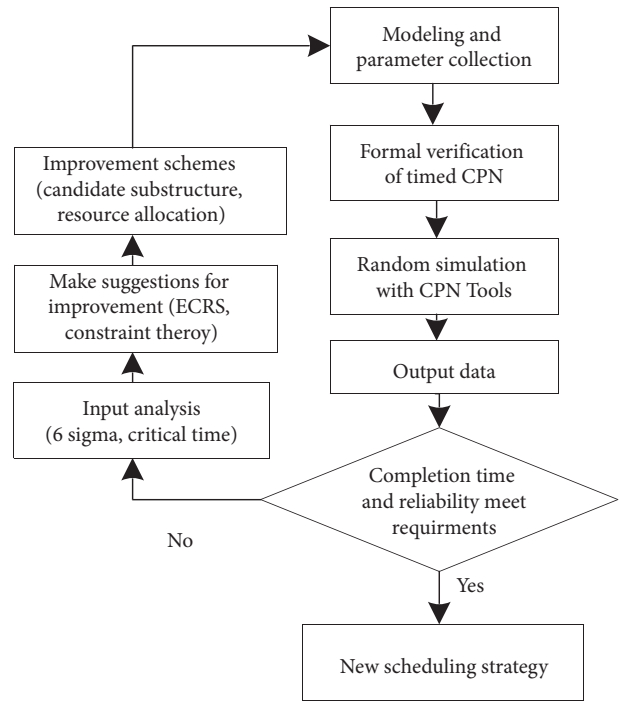


Figure 1: Research process.

according to the original scheme. We then simulate the original model through the CPN Tools and extract the key data in the simulation results; we use the key time analysis method and the 6 sigma criterion to analyze the simulation data; finally, we use the ECRS method and constraint theory to obtain improvements. This process is repeated until the final scheduling strategy and resource allocations that meet the project deadline.

3.1. CPN Theory and Formal Verification. Previously, we illustrated that PN has been successfully used for modeling [19]. Compared with DAG, CPN can express more complicated process, such as or-split selection paths, and-join parallel path, and defined relationships of resource dependency and consumption and the characteristics of space-time networks. Besides, the software support of CPN is comparatively integrated; it supports substructure modeling, color sets of theoretical models, time modeling, and automatic validation.

3.1.1. Petri Net Theory. To facilitate understanding, we have summarized the meaning of the symbols in Table 1.

Various types of extended PN are defined as follows:

(1) A PN is a 5-tuple, PN = (P, T, F, W, M_0) where the PN structure $N = (P, T, F, W)$ without any spec specific initial marking is denoted by N.

(2) A CPN, $R = \{P, T, C, I^+, I^-, M\}$ where an item of $C(s)$ is called a color of s and $C(s)$ is called the color set of s.

(3) TPN was developed based on traditional PN. In TPN (timed Petri net), each transition takes a "real time"

Table 1: PN element interpretation.

PN element	Explanation
P	$P = \{p_1, p_2, \ldots, p_m\}$, a finite set of places
T	$t_1, t_2, \ldots, t_n\}$, a finite set of transitions
F	$F \subseteq (P \times T) \bigcup (T \times P)$, a set of arcs (flow relations)
W	$\longrightarrow$ {1, 2, 3, ...}, a nonempty and finite set of weightings for arcs
M_0	$P \longrightarrow$ {1, 2, 3, ...}, $P \bigcup T \neq \varnothing$, $P \cap T \neq \varnothing$, initial marking
C	A colored function from $P \bigcup T$ to W
I^+, I^-	Forward (backward) incidence matrix of $C \times T$, where $I^+(p,t)$ is a function from $C(p) \times C(t)$ to rmN, where N is the set of nonnegative integers
M	Initial marking of the net which is a vector of P, where $M(p)$ is a function from $C(p)$ to N
f	Firing time function, which assigns the nonnegative (average) firing time $f(t)$ to each transition t of the net
n	The number of classes that model the system dynamics
R	Priority index for different products or jobs
V_i	The set of variables in NHTCPN

to fire (i.e., there is a firing time associated with each transition of a net that determines the duration of the firing of a transition [20]).

A conflict-free TPN is a pair $N_{\text{timed}} = (M, f)$, where M is a conflict-free marked PN, $M = (N, M_0)$, $N_{\text{timed}} \longrightarrow R^+$, and R^+ denotes the set of nonnegative real numbers [21].

(4) A timed CPN (TCPN) is a structure defined as $N = (P, T, F, W, M_0, f, R)$, which is developed based on the traditional PN, CPN, and TPN; hence, it should contain all of the elements of these three existing PNs. Besides that, a new element, priority for different product or job R, is introduced for the new PN in (1). In real cases, different products or different production lines will adjust the processing sequence based on its current significance, such as customer-wanted date or the importance index of the customers.

(5) Hierarchical TCPN (HTCPN) models the dynamic evolution of a system [22]. It is composed of a set of HTCPN subnets. The definition for the HTCPN is given as follows:

(a) An HTCPN is composed of a finite set of HTCPN subnets (i.e., NHTCPN = $\{\text{NHTCPN}_1, \text{HTCPN}_2, \ldots, \text{NHTCPN}_l\}$, where l is the number of classes that model the system dynamics.)

(b) $\text{NHTCPN}_i = \{-\text{A}_i\text{NTCPN}_i, V_i, M_{0i}\}$, where A_i is the name of the subnet and NTCPN_i is a TCPN (following (4))

The definition of the HTCPN subnet is based on the identity structure of a certain network. In a complex hierarchical system, when modeling from a holistic perspective, there may be multiple alternate transitions, and each is associated with a complex HTCPN termed a subnet. Subnets allow a more specific and accurate interpretation of the activities in the replacement transformation. Each subnet can be thought of as a unique integrated HTCPN, but its inputs and outputs are closely tied to other subnets. Therefore, HTCPN can reduce the complexity of modeling and improve readability.

3.1.2. Petri Net Properties. For the verification of the manufacturing process model, some properties and analysis methods of the basic Petri net are used, among which the properties involve reachability, boundedness, activity, and fairness. A common analysis method is to construct the reachability identification map of the Petri net, which can be used to analyze the state of the network system and the sequence of its occurrence in order to know the relevant properties of the network system. Finally, rules based on these properties are specified to achieve the purpose of automatic formal verification. The natural and formal verification theorems are as follows:

Definition 1. Reachability of the Backbone Process. Under the network identifier M_0, the only execution token E is in the starting place of the process. After the limited sequence of transitions, the network identifier arrives, and the only execution token E is located in the last place of the process.

Definition 2. Reachability of Task Subprocess. In contrast to the main process, one or more execution tokens E are in the starting place of the process in the network identifier $M\prime$. After a limited sequence of transitions, the network identifier M'' arrives, and all the execution tokens E are in the ending place of the process. As a result, the subprocess of tasks supports the multithreaded execution, and this subprocess can simulate parallel and serial execution of multiuser, multirole, and multitask.

Definition 3. Activity. After adding a virtual line (transition and the color of the input arc and output arc is the execution token) between the place where the process starts and ends, for any two libraries P1 and P2, if the execution token E is in P1, after a limited transition sequence, E reaches P2. That is, only the CPN of the color set E is examined, at which point it is degraded to the basic Petri network, which is verified by the definition of its activity.

Definition 4. Fairness. After adding a virtual line (transition and the color of the input arc and output arc is the execution token) between the beginning place and the ending place of

the process, after a limited transition sequence, each place has the opportunity to receive an execution token.

Definition 5. Reachable Marked Graph. The reachable identification set $R(M_0)$ of a bounded Petri net is a finite set. $R(M_0)$ can be used as a vertex set, and a direct reachability relationship between the identifiers is used to form a directed graph. This graph is termed a Petri net's reachable marking graph, and it can be used to analyze the state changes and transition occurrence sequences of the network system in order to know the relevant properties of the network system.

Definition 6. Formal Verification Rules for Workflow.

Rule (1): stimulation of changes and changes in multiple sets of color examples to the place follow HTCPN.

Rule (2): in the initial state M_0, the only execution token E is in the starting place of the trunk process; when the trunk process stimulates parallel transitions, the unique execution token copies an execution token for each branch, which is referred to as explicit defined parallelism; when other types of control transitions (serial, select, or loop) are fired, the unique execution token remains unchanged and only behaves as a pass.

Rule (3): for task subprocesses, the operating mechanism of serial, selection, or cyclic control changes is the same as Rule (2). Parallel control changes need to be classified and discussed.

Explicitly defined parallelism between a single execution token, as in Rule (2); the implicit parallelism between multiple execution tokens is a single branch of the CPN graphic definition repeatedly. With the multiple execution tokens, the same transitions are triggered multiple times simultaneously. Dynamic, on-demand simultaneous creation of multiple execution tokens is involved, creating tasks in parallel; the serialization of multiple execution tokens is a single branch of the CPN graphic definition. With multiple execution tokens, the same transitions are triggered in a controlled manner multiple times. This approach creates multiple execution tokens dynamically, on demand, and serially.

3.1.3. Steps for Modeling Method with HTCPN

Step 1. Outline the overview structure of the object-distributed manufacturing network, and transform all the elements into the above CPN elements to build up first-level PN.

Step 2. Further define the subnet for the first-level HTCPN.

Step 3. Check whether the subnet meets the above definition and verification rules, and the subnets are always the substitution for transition element of the last level. If it meets the requirements, go to Step 5; otherwise, go to Step 4.

Step 4. Modify the PN model structure to make it conform to all definitions and rules, and go to Step 3.

Step 5. Repeat Step 2 until the transition elements are not required for further development, which means that the basic elements in a distributed manufacturing network have been modeled.

Step 6. Ensure that all of the links between the levels are consistent through the entire HTCPN.

3.2. Optimization Mathematical Model. According to the steps of HCPN modeling in the previous section, if the workflow is effective, it must meet the definition and rules of PN. Next, we need to calculate the workflow completion time of the current iteration under the condition that the workflow is valid (meets the latest deadline and is reliable), and we express the calculation process by means of a mathematical model.

The variable interpretation is shown in Table 2.

The mathematical model is as follows:

$$\operatorname{Min}\left(T^{\mathrm{SD}}\right), \tag{1}$$

$$T^{\mathrm{Avg}} \le D, \tag{2}$$

$$T^{s}_{ijk} \ge \mathrm{T}^{\mathrm{ESD}}_{ijk}; \quad i = (0, 1, 2, \ldots, n),\ j = \left(1, 2, 3, \ldots, \mu_i\right),\ k = \left(1, 2, 3, \ldots, \sigma_{ij}\right), \tag{3}$$

$$T^{f}_{ijk} = T^{s}_{ijk} + T^{e}_{ijk}; \quad i = (0, 1, 2, \ldots, n),\ j = \left(1, 2, 3, \ldots, \mu_i\right),\ k = \left(1, 2, 3, \ldots, \sigma_{ij}\right), \tag{4}$$

$$T^{f}_{ijk} \le T^{\mathrm{LED}}_{ijk}; \quad i = (0, 1, 2, \ldots, n),\ j = \left(1, 2, 3, \ldots, \mu_i\right), k = \left(1, 2, 3, \ldots, \sigma_{ij}\right), \tag{5}$$

$$T^{i}_{\mathrm{cirtical}} = \max_{\substack{j\epsilon\left(1,2,\ldots,\mu_i\right)\\ k\epsilon\left(1,2,\ldots,\sigma_{ij}\right)}} \left(T^{f}_{ijk}\right), \quad i = (1, 2, \ldots, n),\ j = \left(1, 2, 3, \ldots, \mu_i\right),\ k = \left(1, 2, 3, \ldots, \sigma_{ij}\right), \tag{6}$$

Table 2: Mathematical variable interpretation.

Symbol	Description
n	The number of workflows
W	The set of workflows $W = \{W_1, W_2, \ldots, W_n\}$
W_i	The ith workflow in the workflow set
μ_i	The number of substructures in the workflow W_i
V_i	The set of substructures in workflow W_i, $V_i = \{V_{i1}, V_{i1}, \ldots, V_{iu_i}\}$
V_{ij}	The jth substructure of workflow W_i
σ_{ij}	The number of tasks of substructure V_{ij}
U_{ij}	The task set of substructureV_{ij}, $U_{ij} = \{U_{ij1}, U_{ij2}, \ldots, U_{ij\sigma_{ij}}\}$
U_{ijk}	The kth task in the substructure U_{ij}
P_{ijk}	Predecessor set of task U_{ijk}
D	Deadline for the project
m	The number of resource types
$\overrightarrow{R}^f(t)$	The resource vector owned at the current time t $\overrightarrow{R}^f(t) = \{R_1^f(t), R_2^f(t), \ldots, R_m^f(t)\}$
$\overrightarrow{R}^r(i,j,k)$	Resource vector required for task U_{ij} execution $\overrightarrow{R}^r(i,j,k) = \{R_1^r(i,j,k), R_2^r(i,j,k), \ldots, R_m^r(i,j,k)\}$
T_{ijk}^e	Time required for task U_{ijk} to execute
T_{ijk}^s	Time when task U_{ijk} started to execute
T_{ijk}^f	Time when task U_{ijk} execution ends
T_{ijk}^{ESD}	The earliest start time of the task U_{ijk} (calculated according to the project start time)
T_{ijk}^{LED}	The latest end time of the task U_{ijk} (calculated according to the project deadline)
$T_{\mathrm{criticalw}}^i$	Time to complete the critical path of workflow W_i
T^{Avg}	Average completion time of all workflows
T^{SD}	Standard deviation of completion time

$$x_{ijk}(t) = \begin{cases} 1, & \text{Task } U_{ijk} \text{ executed at time } t, \\ 0\,\mathrm{c} & \text{Otherwise,} \end{cases} \quad i = (0, 1, 2, \ldots, n),\ j = (1, 2, 3, \ldots, \mu_i),\ k = (1, 2, 3, \ldots, \sigma_{ij}), \tag{7}$$

$$\sum_{i=1}^{n} \sum_{j=1}^{\mu_i} \sum_{k=1}^{\sigma_{ij}} (t) \cdot \overrightarrow{R}^r(i,j,k) \le \overrightarrow{R}^f(t), \tag{8}$$

$$\mathrm{T}_{ijk}^s \ge \mathrm{Max}_{k' \epsilon P_{ijk}} \left(\mathrm{T}_{ijk}^f\right), \quad i = (0, 1, 2, \ldots, n),\ j = (1, 2, 3, \ldots, \mu_i),\ k = (1, 2, 3, \ldots, \sigma_{ij}), \tag{9}$$

$$T^{\mathrm{Avg}} = \frac{\sum_{i=1}^{n} T_{\mathrm{critical}}^i}{n}, \tag{10}$$

$$T^{\mathrm{SD}} = \sqrt{\frac{\sum_{i=1}^{n} \left(T_{\mathrm{critical}}^i - T^{\mathrm{Avg}}\right)^2}{n}}. \tag{11}$$

Equation (1) represents the objective function of the mathematical model. The purpose of using the minimized standard deviation as the target value is to provide a basis for the reliability analysis in the following section; equation (2) indicates that the average value of all workflow simulation results must be less than the project deadline; equation (4) is used to calculate the completion time of the task, and equations (3) and (5) define the time window constraint of the task U_{ijk}; that is, the start time of the task is later than the earliest start time of the task and the end time. It must be earlier than the latest end time of the task. Among them, T_{ijk}^{ESD} and T_{ijk}^{LED} can be calculated according to the start time and deadline of the space project, and we will explain this in detail in the experimental analysis section. In equation (6), we use the concept of completion time of the critical substructure consistent with the following sections to represent the completion time of the entire workflow; in equation (7), $x_{ijk}(t)$ represents the decision variable, and when the task U_{ijk} is executed at time t, $x_{ijk}(t)$ is equal to 1; otherwise, $x_{ijk}(t)$ is equal to 0; the resource limit is defined in equation (8) and represents the resources required for task execution at any time that cannot exceed the resources currently owned; equation (9) expresses the timing constraint relationship between tasks—the current task can only be executed after the execution of all previous tasks is complete; equations (10) and (11) represent the calculation process of the average and standard deviation of workflow completion time, respectively.

3.3. Analysis Methods. In this section, we will introduce some methods that will be used in the simulation analysis process.

3.3.1. 6 Sigma System. To ensure the reliability of the manufacturing process, the 6 sigma system was introduced to the simulation iteration.

The 6 sigma model is shown in Figure 2. The USL and LSL are the limits of this system. If the result is normally distributed, the distribution diagram is within the acceptable range of plus or minus three sigma, which is the standard deviation [23], and the results located outside the area will be 0.135% [24]. Hence, if the company wishes to maintain the service level target at 99.865%, the data derived from the case CPN simulation require a new workflow; the 6 sigma system was introduced to the simulation iteration to control the completion time before the project deadline. 6 sigma has been demonstrated to be an effective management framework and methodology for improving operation workflows in industry [25, 26]; various quality systems in engineering and service industry use this systematic approach to implement quality standards for process control and monitoring. Our case study of space instruments, hence, makes use of this strategy to improve the detailed operations, from simulation to implementation of the quality management system [27, 28]: it follows DMAIC (Define, Measure, Analyze, Improve, and Control). Define, Measure, and Analysis are carried out through simulations of workflow changes described in Section 6.1, and Improve and Control are performed with elimination and simplification following Control of critical resources and allocation of assemble, rework, and inspection tasks, as illustrated in Section 6.2.

3.3.2. Critical Time Analysis. During the initial design phase of the manufacturing process, each task is assigned to the theoretically earliest start date (TESD), which is the earliest time that a task can begin without considering resource constraints. The start time of each task cannot be earlier than TESD because some tasks need to wait for the predecessor to complete execution. The TESD of each task can be calculated according to equation (12), and the LT_n represents the execution time of task n:

$$\mathrm{TESD}_{n+1} = \mathrm{TESD}_n + \mathrm{LT}_n. \tag{12}$$

The latest start date (LSD) is another important data point to analyze. The LSD of the task is calculated forward according to the deadline for the entire process. If the task cannot be initiated before the latest start time, the entire system may not be completed within the specified deadline. The LSD of each task can be calculated according to the following equation:

$$\mathrm{LSD}_n = \mathrm{LSD}_{n+1} - \mathrm{LT}_n. \tag{13}$$

Through the TESD and LSD of each task, we can calculate the time window of each task, which is LSD-TESD. If there is a task in the simulation results that cannot be completed within the time window, the task is the key object to be analyzed. The specific situation will be introduced in detail in Section 5.

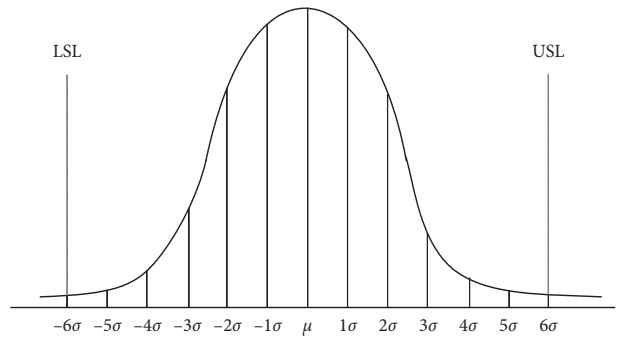


Figure 2: Six sigma.

3.3.3. ECRS Method. The ECRS method is one of the most fundamental activity processing tools in industrial engineering [18]. The letter *E* stands for elimination of unnecessary work, C for combination of operations, *R* for activity sequence rearrangement, and *S* for necessary work simplifications. These four principles have been implemented in many industries. The ECRS method has been used in ice cream manufacturing processes [29], electronic manufacturing industries [30], etc.

The ECRS method will be the base principle for this improvement proposal. For every critical issue, this dissertation will attempt was proposed by Dr. Goldratt in 1990 [29]. This theory was developed for optimized production technology, focusing on identifying the constraints inhibiting the successful processing of the workflow and then solving them. The theory of constraints lists eight rules for manufacturing process planning:

(a) Balance logistics or material instead of balance capacity.

(b) Noncritical resource utilization is decided by the system constraints instead of its own potential capacity.

(c) High resource utilization does not necessarily mean that the system is working in an efficient manner.

(d) Delays in critical path usually mean delays to the entire system.

(e) Saving time or resources in noncritical paths often has no positive impact on the system.

(f) Critical paths normally will decide most properties of the system, for example, the system lead time, system productivity, and inventory.

(g) Batch size should vary with time.

(h) Priority should be designed based on the system constraints. Early finished tasks are not resulted from the initial design but the processing of the plan.

Those eight rules mention a particularly important improvement criterion and the critical path, and provide substantial inspiration for the improvement of our workflow

structure. We will elaborate upon specific improvement measures in the experimental result analysis section.

4. Case Modeling and Simulation

4.1. Case Description. In the case of application, the space mission is to explore Phobos, which is one of the two moons that orbit around Mars. The Phobos-Grunt mission planned to land on Phobos to retrieve samples for the purpose of scientific research and also to serve as an illustration for future sample-retrieving space missions [31]. The mission failed because the spacecraft never managed to leave the earth's orbit [32]. The second attempt of the Phobos-Grunt mission is planned in 2025 with minor amendments, which makes the research focus of this paper not only illustrative, but also practical. The architecture of the Phobos-Grunt soil preparation system (SOPSYS) to be landed on Phobos is shown in Figure 3 [10]. The actuator E will first deliver the sample particles inside the spacecraft for analysis and then deliver the debris outside the spacecraft into space. The study focus is Part E's manufacturing process.

The delivery unit E consists of five parts, and the code name for the parts in its manufacturing process is PART0030 (delivery motor with encoder), PART0031 (sieve encoder PCB), PART0032 (sieve encoder receiver), PART0033 (sieve encoder LED), and PART0034 (delivery chamber). One subpart of CPN0012 (delivery encoder) is assembled with PART0031 and PART0032. The delivery units deliver the sample particles into the spacecraft and deliver debris into space.

There are five parts that must be purchased or manufactured throughout the process; the sequence is shown in Figure 4. There are 5 engineers, 4 operators, and 1 inspector in this process. The entire process begins with the initial design of CPN0005 (the Phobos-Grunt SOPSYS delivery unit). After the CPN0005 is designed, 4 tasks can then start to proceed simultaneously: the detailed specification of PART0030 and PART0033, the design and budgeting of CPN0012, and the design, budgeting, and test equipment design of PART0034. Purchasing PART0030 and PART0033 comes after the detail specification, and purchasing PART0031 and PART0032 comes after the design of CPN0012. PART0034 is the only part that needs to be in-house manufactured. The manufacturing of PART0034 will begin immediately after its design phase. After every purchasing, manufacturing, and assembly process, there will be an inspection process to ensure its quality. Hence, when PART0030, PART0031, PART0032, and PART0033 are received and PART0034 is manufactured, they need to be inspected to proceed to the later assembly process. After PART0031 and PART0032 are inspected, CPN0012 will then be assembled, and another inspection comes after it. Finally, PART0030, CPN0012, PART0033, and PART0034 will be assembled together into CPN0005, which will be tested and inspected. The network flow of the entire original manufacturing process and required resources are shown in Table 3.

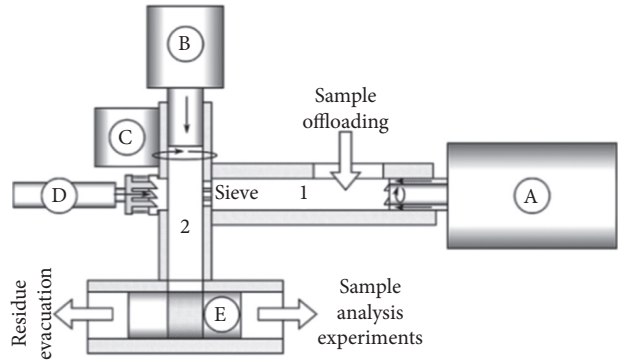


Figure 3: Architecture of SOPSYS.

4.2. Mapping Workflow Elements into CPN. In this section, we will describe the basic workflow structure of the HCPN [33], thus formalizing the SOPSYS process model. First, we mapped between the Phobos-Grunt SOPSYS delivery unit manufacturing process case and CPN, as shown in Table 4.

According to the description of (1) in the first part of the Methodology section, $P = \{p_1, p_2, ..., p_m\}$, with a finite number of places (nodes). Places in this CPN can be imagined as a warehouse between different tasks that holds the material along the manufacturing process and the staff pool that provides human resources. Two different places do not necessarily have to be two different physical locations because places are defined by the tokens they are holding.

$T = \{t_1, t_2, ..., t_n\}$ is a finite transition that represents the dynamic process of advancing from one node to the next, due to certain conditions or rules being met. By defining the tasks as a transition, we can naturally model the resources and products as tokens in the CPN. Tokens in the CPN can represent the resources and products consumed and produced along the PN. For example, for the task PART0032, the inspection requires one inspector and one uninspected PART0032 to produce an inspected PART0032. Two tokens are the prerequisite for this transition, one represents the inspector, and one represents the uninspected PART0032. In the firing process of this transition, these two tokens are consumed, and two new tokens are generated. Consequently, one token representing the inspected PART0032 will be ready for its next transition, and another token representing an idle inspector will be generated for its next transition.

In the SOPSYS model, different colors represent various types of things; therefore, the color can be the thing it stands for. In this case modeling, there are four colors in total. Each color represents a type of resource or token. The three colors are color engineer, color operator, and color inspector.

$f = (\text{REFT} : \text{T} \longrightarrow \text{IN}, \text{RLFT}: \text{T} \longrightarrow \text{IN} \cup \infty)$ represents functions describing the relative earliest and the relative latest firing times of the transitions, where clearly $\text{REFT}(t) \leq \text{RLFT}(t)$ for each $t \in T$, where IN is the set of naturals (including 0). In the context of SOPSYS, the relative earliest firing time and the relative latest firing time of the transitions are firing domains of enabled transitions. The definition of REFT and RLFT differs on transitions with different colored types [14].

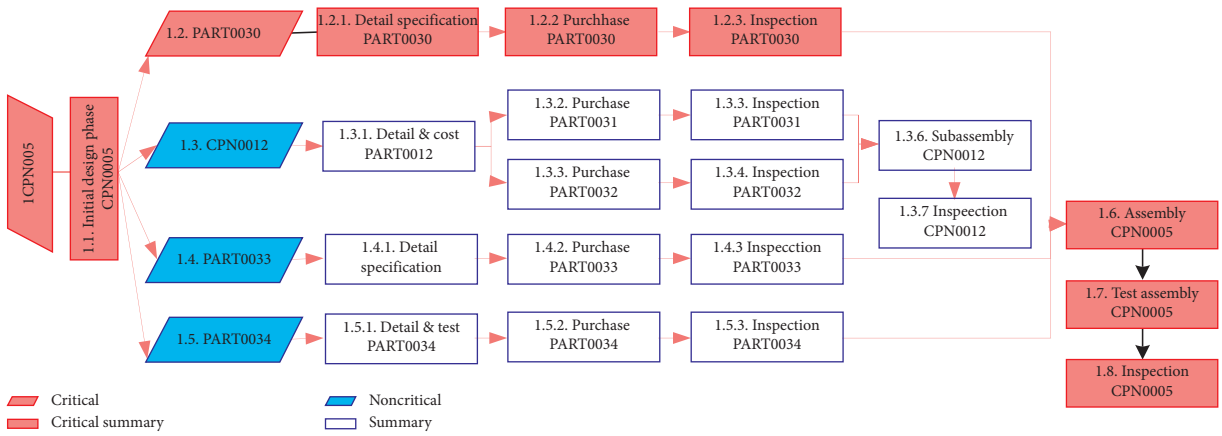


Figure 4: Manufacturing process.

Table 3: Manufacturing process and required resources.

ID	Task name	Duration (days)	Predecessor	Resource name (number)
1	1. CPN0005	100		
2	1.1. Initial design phase CPN0005	40		Designer (2)
3	1.2. PART0030	38		
4	1.2.1. Detail specification PART0030	5	2	Designer (2)
5	1.2.2. Purchase PART0030	30	4	
6	1.2.3. Inspection PART0030	3	5	Inspector (1)
7	1.3. CPN0012	34		
8	1.3.1. Design & cost CPN0012	7	2	Designer (2)
9	1.3.2. Purchase PART0031	21	8	
10	1.3.3. Purchase PART003	5	8	
11	1.3.4. Inspection PART0031	1	9	Inspector (1)
12	1.3.5. Inspection PART0032	3	10	Inspector (1)
13	1.3.6. subassembly CPN0012	2	11, 12	Operator (1)
14	1.3.7. Inspection CPN0012	3	13	Inspector (1)
15	1.4. PART0033	23		
16	1.4.1. Detail specification PART0033	5	2	Designer (2)
17	1.4.2. Purchase PART0033	15	16	
18	1.4.3 Inspection PART0033	3	7	Inspector (1)
19	1.5. PART0034	21		
20	1.5.1. Design & test CPN0034	8	2	Designer (2)
21	1.5.2. Manufacturing PART0034	10	20	Operator (2)
22	1.5.3. Inspection PART0034	3	21	Inspector
23	1.6. Assembly CPN0005	8	16, 14, 18, 22	Operator (4)
24	1.7. Test CPN0005	7	23	Inspector (4)
26	1.8. Inspection CPN0005	7	24	Inspector

Table 4: Mapping between case and CPN.

Workflow element	CPN element
Tasks	Transitions
Resources and products	Tokens
Types of resources and products	Colors
Warehouse and staff pool	Places

HTCPN = (A_i, $NTCPN_i$, V_i, M_{0_i}) is an HCPN model that characterizes the hierarchical model of the SOPSYS. This paper will construct a CPN model using a bottom-up approach. The subprocess will be modeled as a subnet, and the subnet will be drawn as a transformation in the CPN overview [34]. In this case, the HCPN does not apply to resources being modeled into the CPN; therefore, it is difficult to find a single port subnet. However, developing a model using a bottom-up approach can improve the readability of the CPN. The first step is to build a simple CPN for different subprocesses.

There are five parts in total. Two parts have the same process and can be described in Figure 5.

As shown in Figure 5, PART0030 and PART0033 both begin from a place name P1 (different places in the final CPN). The first transition is specification, which consumes tokens from P1 and Engineer. Token in P1 stands for the status where PART0030 or PART0033 is ready to be

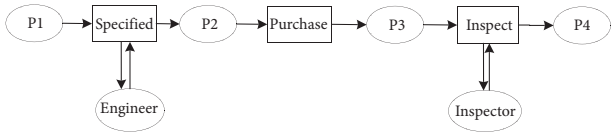


Figure 5: Subprocess of PART0030 and PART0033.

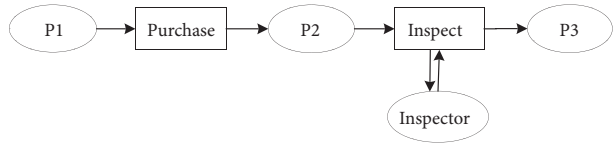


Figure 6: Subprocess CPN for PART0031 and PART0032.

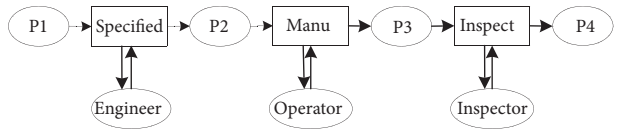


Figure 7: Subprocesses CPN of PART0034.

specified. Token in Engineer stands for engineers that are available for the task. After the transition specified is fired, new tokens will be generated into P2 and Engineer. Newly generated tokens in P2 have the same color set as the tokens consumed from P1, which stands for the status of the product. Newly generated tokens in Engineer have the same color set as the tokens in Engineer, which stands for available engineers. Similarly, transition inspections consume product status color tokens from P3 inspector color token from place Inspector and generate product status color token into P4 and inspector color token back to place Inspector.

Two other parts also share the same process, which is shown in Figure 6.

PART0031 and PART0032 begin immediately after the design of CPN0012, and therefore, the only difference here is that this process does not have a specification transition.

The only special process is PART0034. The process of PART0034 is shown in Figure 7.

As shown in Figure 7, PART0034 is in-house manufactured, which requires operators in the manufacturing process. The operator color token will be consumed from the place Operator, and the product status color token will be consumed from P2. New operator color tokens will be generated back to the place Operator, and new product status color tokens will be generated into P3 for further processing.

The process of CPN0012 can be modeled as follows.

Figure 8 shows the subnet for the process of CPN0012. There are two aspects to be noticed in the subprocess of CPN0012. The first is that P2 now provides tokens for both PART0031 and PART0032, and these two transitions can be fired concurrently. Thus, P2 must have two tokens in place to trigger both transitions. The next aspect is that two places of Inspector in Figure 7 are the same and contain only a limited number of inspectors. In this case, there is only one inspector available.

There are four places at the beginning of the middle stream to differentiate the four processes; otherwise, one of the transitions might be fired multiple times and other transitions will not be fired at all. If the tokens standing for the products status are divided into four types, which can represent four different subparts, then a more readable CPN can be established, as in Figure 9. The transition subpart is a complicated subnet containing all processes of the four parts. It is plotted as a transition to increase the readability of the CPN, but it is not a single port subnet, and therefore, it cannot be simulated in this manner.

4.3. CPN Tools Modeling. CPN Tools is a software tool for editing, simulating, and analyzing CPNs [35]. CPN Tools can simulate the performance of the model in real time. The declaration in CPN Tools defines the data type in the CPN model [36]. The declaration of the CPN Tools model is shown at the bottom right of Figure 10. When modeling a CPN into CPN Tools, the structure of the net must be modified. It is necessary to model the CPN with the computer logic. The structural modification includes the construction of the following parameters: real-time element, setting place color set, rework structure, and construction of output transition.

Real-Time Elements. The real-time element is integrated into the CPN Tools model through adding a time lag to transitions. Figure 10 shows how to achieve this.

The syntax for implementing real-time is @+(time). Figures 4–10 indicate that transition Inspect will take 7 units of real time to fire. If the timed token consumed in this transition is (elements in a color set) @t, the newly generated token would be (elements in a color set) @($t+7$).

Define Color Groups. A place can only hold certain types of tokens. In the left-hand side of Figure 10, the text "inspector" in the lower-right corner indicates its color setting. If there are multiple types of tokens in one place, their color set should be specifically defined as a list of token types. If a place holds more than one type of token, its color set should be specially defined as lists of token types. A list is a tuple containing variables from different color sets. For example, in this case, if a place represents a pool of staff for all three types of worker (engineers, operators, and inspectors), the elements in the color set here are (e, O, and I).

As shown in the left-hand side of Figure 10, the task before an inspection will randomly generate an integer token with a value ranging from 1 to 100. The model assumes that all these tasks will have a scrap rate or rework rate of 5%. Thus, the code segment of the inspection transition will use value 5 as the differentiator of a good product and a bad product. If the value is less than or equal to 5, then a token with color bool value false will be generated into the next place. This will not trigger anything further ahead. In the meantime, another token with color value true will be generated into the rework place. This token will trigger the rework transition, which will send the M token back before the task before inspection. If the randomly generated value is greater than 5, then the process moves forward as expected.

An additional transition must established in the CPN to export the performance data. The transition is shown at the

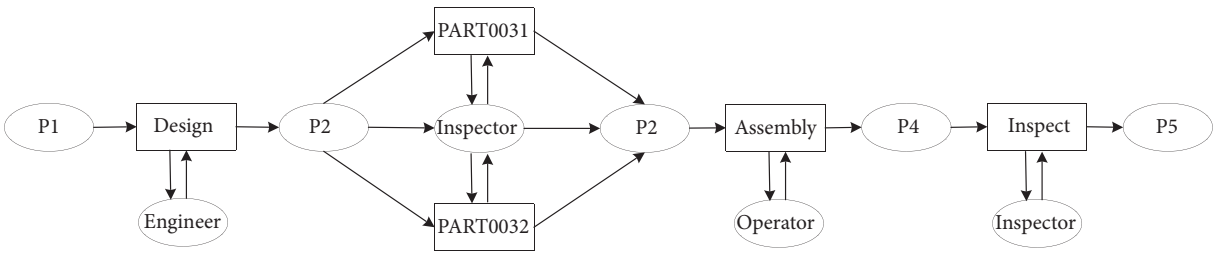


Figure 8: Subprocess CPN of CPN0012.

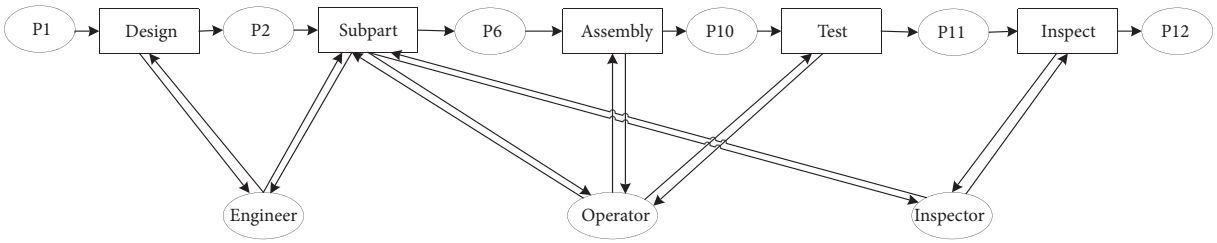


Figure 9: CPN model for the whole process.

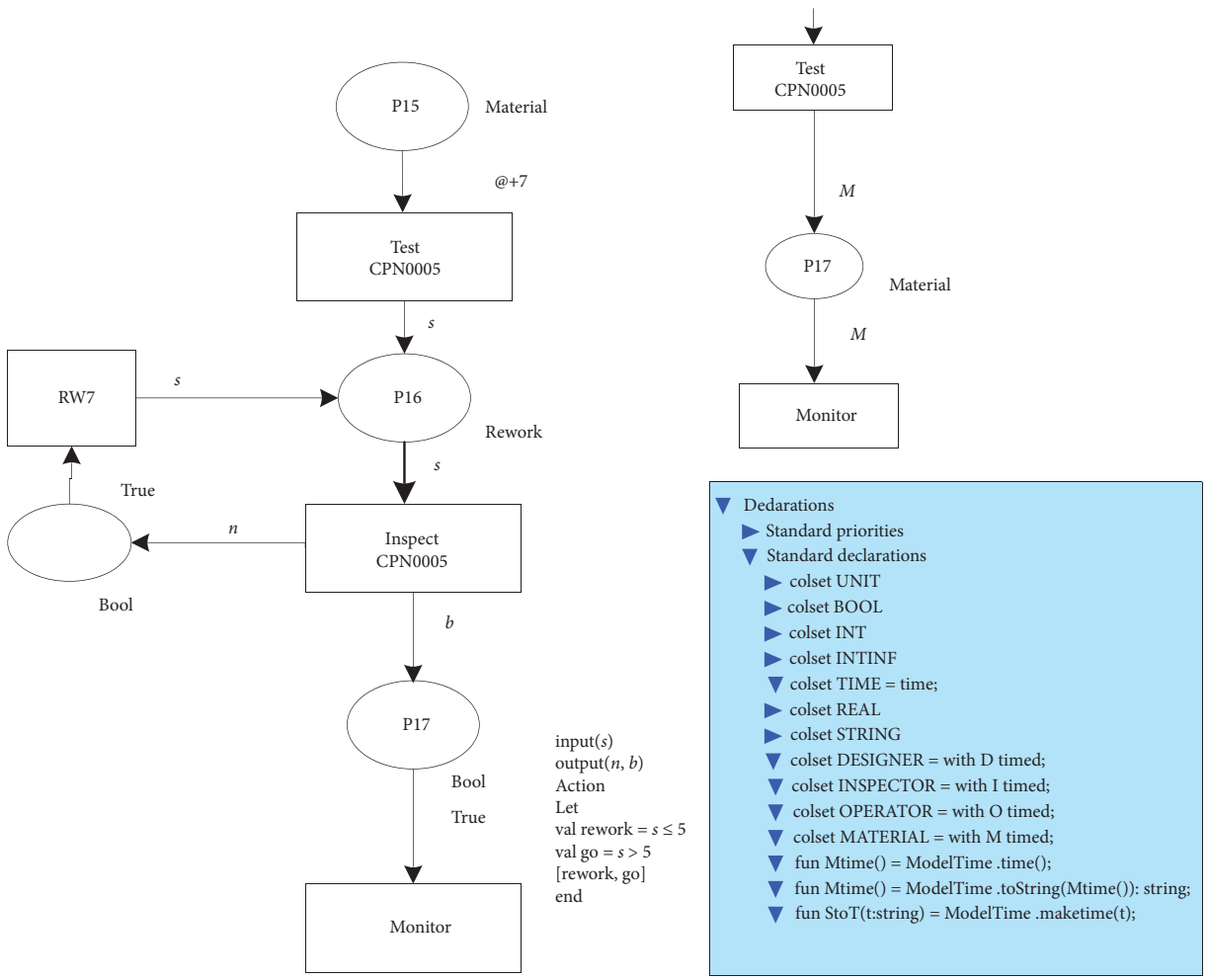


Figure 10: Rework structure and output transition.

top right-hand side of Figure 10. After CPN0005, the delivery unit is inspected and confirmed qualified; then, the token in the final place (indicating a finished delivery unit) will be consumed in another transition referred to as "monitor." This monitor transition will consume the token and report the time stamp attached to the token, which is the total lead time.

In summary, we model the mainstream process as CPN and describe the rework structure, and the inspection has been modeled as a normal transition, which will drive the process forward. Furthermore, according to Definitions 1–6, the model meets all formal verification rules. The complete model is given in Figure 11.

5. Simulation Result and Analysis

5.1. Simulation Result. The data type has already been introduced in the previous section, and the model will be simulated 10,000 times to ensure all possible conditions could occur. The rework rate for each task before an inspection is 5%, which is one out of 20. These types of tasks appear three times at most in one work stream. Three tasks in the CPN0012 workflow stream need to be inspected. Hence, if the rework situation is uniformly distributed, the likelihood of it occurring three times in a row is $20 * 20 * 20 = 8000$. Simulating the CPN 10,000 times will theoretically generate the results for most possible conditions.

After 10,000 repeats of the simulation, the average lead time is 118.06 days. The best possible lead time is 109 days where resources are perfectly allocated and all tasks have been approved by inspection for just one time (see Tables 4 and 5 for detailed indicators). However, the deadline for this project is 100 days; therefore, it cannot meet the demands, and the process should be improved.

5.2. Result Analysis

5.2.1. Analysis of Critical Time. Based on the simulation results, we summarize the data on the start time as shown in Table 5. Before visualizing all the summed data, the actual earliest starting date (AESD), theoretical earliest starting date (TESD), and latest starting date (LSD) should be compared. The DELAY value is equal to AESD-TESD, indicating how much the actual start time is later than TESD, which may be due to resource constraints that are not theoretically earliest. The FAIL value is equal to AESD-LSD. When the FAIL value is greater than 0 (marked in red), it indicates that the actual start time of the task is greater than the latest start time of the task, which will inevitably cause the completion time of the entire system to exceed the specified time range. WINDOW is equal to LSE-TESD, which represents the buffer time of the task (i.e., the task starts on any day within the time window, and the entire manufacturing system will not exceed the specified completion time). If the DELAY value is greater than WINDOW, the manufacturing process cannot be completed on time, which is the same as the judgment using the FAIL value.

After comparing and listing all these data, we analyze them visually, as shown in Figure 12.

The last three tasks in the process have the same FAIL value because these three tasks are a simple one-stream process. Thus, it is the previous tasks that result in the positive FAIL value.

I0030 is an interesting task. The data show that I0033 has a FAIL value of 1, which means that this task can only begin one day after being scheduled. Its preceding resource-required task is S 0030. The data show that the FAIL value for S0030 is 0, which means that it is not S0030s fault that I0030 is late. Therefore, the only possible cause of the lateness of I0033 should be the limited Inspector resource.

I0012 is another failed task. The logic for analyzing task I0012 is the same as for I0033. The previous resource-required task has a FAIL value of −4 and the shortage of Inspector for task I0012 will, therefore, be solely responsible for the failure. Thus, critical issue one is caused by the shortage of inspectors.

Five of the tasks have no window at all. This is not acceptable, particularly when there are two tasks that might require rework. If any of the no-window tasks are reworked, the total service level will at least decrease to 95%, which is the service level for this task.

We have observed that most tasks have a delay time more than their windows. This could be due to many reasons. Consider the logic of the CPN; there are only three possible reasons for this: rework, limited resources, and bad workflow structure design. Yet, if we examine in further detail, tasks involving engineers generally have a delay time similar to or less than the window. Most of the delay problems occur for the inspectors. Thus, the most important resource shortage that causes this critical issue is also the shortage of inspectors.

5.2.2. Analysis about Reliability. To analyze reliability, we have calculated the following data, as shown in Table 6. We summarize the standard deviation (STD) and the average start time (AVG) of the simulation results, where FAIL is the difference between AVG and LSD, and $\text{FAIL} + 3 * \text{STD}$ is calculated based on the 6 sigma system.

Figure 13 lists a couple of critical issues. The 6 sigma system was introduced in the previous section. If the service level is set to 99.865%, the value of $\text{FAIL} + 3 * \text{STD}$ must be negative. Unfortunately, there are only four tasks that reach the standard.

We have observed that with the exception of I0034 and I0032, almost all inspection tasks have a relatively large standard deviation. Because inspection work is relatively short and predecessors of inspection tasks for different components may not end at the same time, the shortage of inspectors does not cause all inspection tasks to have a large standard deviation, and this is because of potential rework. I0032 has a much smaller standard deviation because the purchase process is rapid. I0034 is a similar case.

6. Improvement Strategy

Here, we will use the ECRS method and theory of constraints that was introduced in the Methodology section to propose improvements and suggest new scheduling strategies.

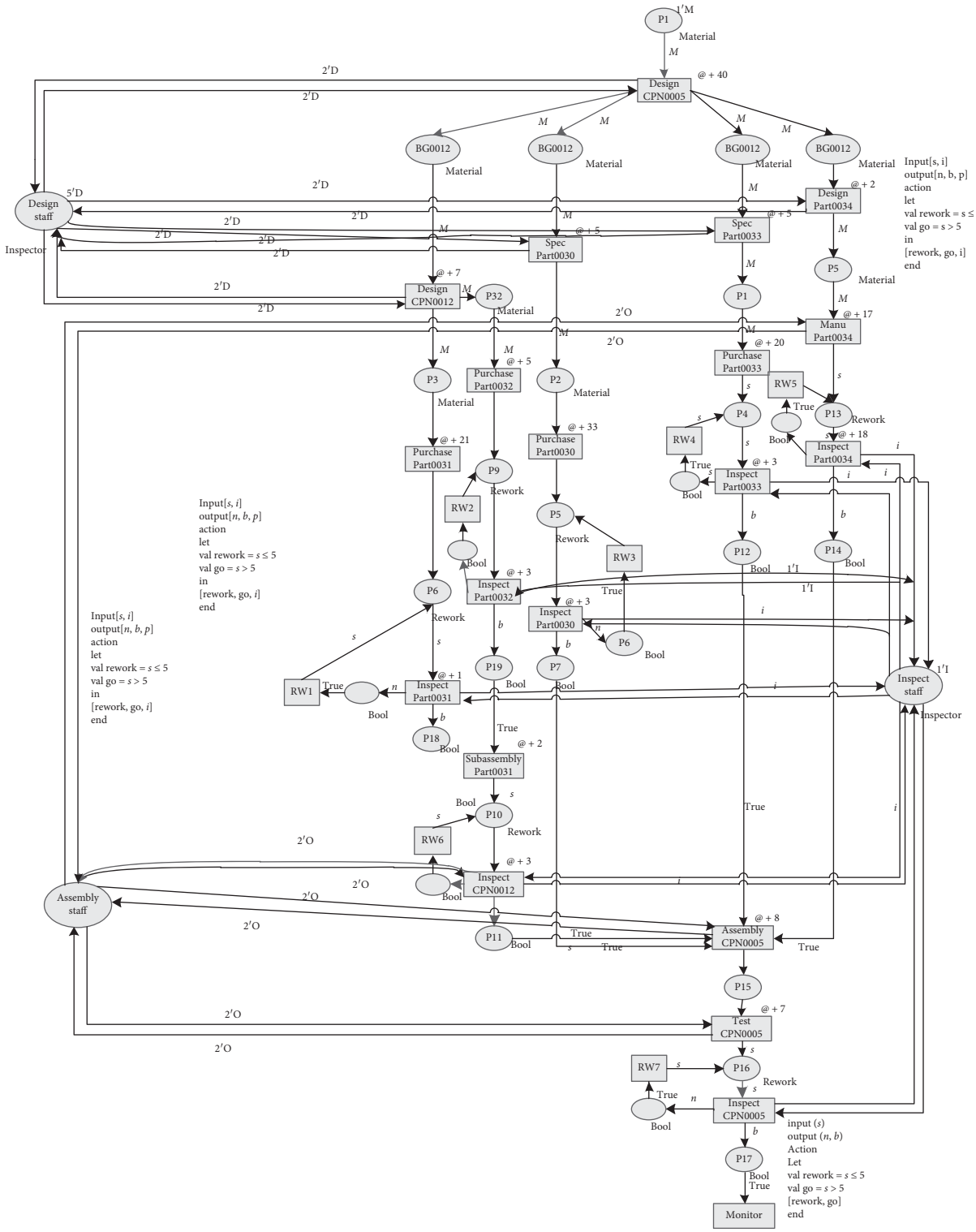


Figure 11: Original CPN model.

6.1. Workflow Structure. According to the theory of constraints, the critical path should first be identified, and then the entire system should be analyzed step by step. The stream of PART0030 is the critical path where all the tasks in it have no window at all. The stream of CPN0012 also has a relatively small window, which could be focused on after the PART0030 stream. PART0033 and PART0034 have a relatively large window, which means that their priorities are in last place.

The workflow structure of PART0030 must be changed because the rework rate will directly lead to the total process reliability being less than 95%. This can be overcome

Table 5: Data summary.

Task	AESD	TESD	LSD	DELAY	FAIL	WINDOW
D0012	40	40	44	0	−4	4
D0034	40	40	57	0	−17	17
S0030	40	40	40	0	0	0
S0033	40	40	55	0	−15	15
A0005	87	78	78	9	9	0
A0012	69	69	73	0	−4	4
M0034	42	42	65	0	−23	23
T0005	95	86	86	9	9	0
I0005	102	93	93	9	9	0
I0012	81	71	75	10	6	4
I0030	77	75	75	2	2	0
I0031	68	68	72	0	−4	4
I0032	52	52	70	0	−18	18
I0033	65	65	75	0	−10	10
I0034	59	59	75	0	−16	16

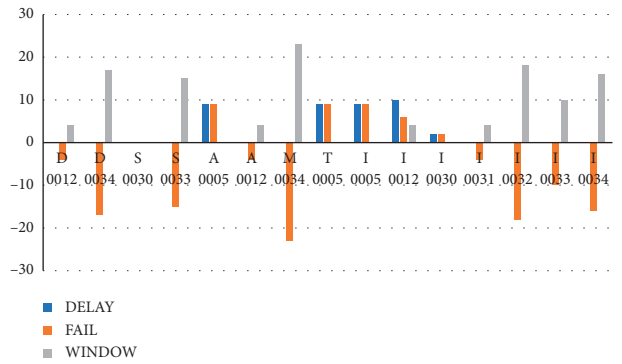


Figure 12: ESD comparison.

Table 6: Reliability data.

Task	AVG	LSD	FAIL	STD	6 sigma
D0012	42.01	44	−1.99	2.24	4.73
D0034	42.8	57	−14.2	2.89	−5.53
S0030	42.38	40	2.38	2.76	10.66
S0033	42.3	55	−12.7	2.75	−4.45
A0005	95.28	78	17.28	10.56	48.96
A0012	84.55	73	11.55	7.06	32.73
M0034	45.69	65	−19.31	5.03	−4.22
T0005	103.67	86	17.67	10.68	49.71
I0005	110.67	93	17.67	10.68	49.71
I0012	88.23	75	13.23	5.97	31.14
I0030	83.72	75	8.72	5.1	24.02
I0031	82.7	72	10.7	5.36	26.78
I0032	54.58	70	−15.42	3.54	−4.8
I0033	80.85	75	5.85	6.39	25.02
I0034	63.06	75	−11.94	5.12	3.42

through buffer times using the ECRS method, and one of which is to change the strategic objective in sourcing and, thus, alter the current suppliers. Suppliers who can deliver the product sooner are preferred. Several models, such as Kraljic's model [37] and the strategic SRM model [38], can support this process. Yet, most of the time, a shorter lead time will lead to worse quality, which will increase the rework rate. Another way to reduce purchase time is to

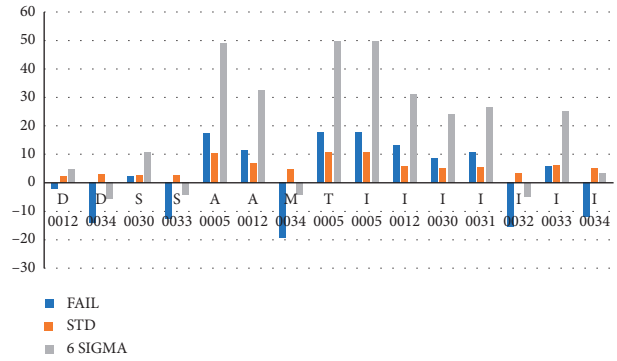


Figure 13: 6 sigma analysis.

provide technical support, such as via engineers or machines for the supplier. Studies have demonstrated that providing technical and resource support can improve the supplier performance in terms of quality and lead time [37]. Thus, engineers can be sent to the supplier to help develop and make PART0030. This approach can also eliminate the inspection task in that engineers that are assisting the purchasing process can help to ensure product quality.

6.2. Assemble, Rework, and Inspection. The appropriate strategy for eliminating rework has already been presented in the previous section. One of the best ways to mediate the negative impact of rework is to find a way to eliminate it. The best possible approach is to send engineers to suppliers to assist with the production process. This strategy has three theoretical positive impacts. The following decisions can be made based on the ECRS method.

One is that the incoming products do not need to be inspected anymore because engineers assisting the production process can do the job instead. The other positive impact is that there will not be any rework risk at all because the production is under inspection for the entire time.

An alternative is to have inspectors doing inspection along the assembly and test tasks. These two tasks might need more time for completion because of quality control, but the risk of large-scale rework is eliminated. Then, the task assembly and test will need a longer and more variable time for completion.

With the improvement strategy proposed, the new model is then simulated, and the execution time of each task and the number of various personnel is randomly generated. During this process, the 6 sigma system is still used as a neighborhood search rule to generate higher-quality candidate solutions to determine the manufacturing time and number of people for each task. Finally, the simulation results of the new manufacturing process are compared with the original model, and the results are as follows.

6.3. Result Comparison. According to the improvement suggestions, we have established a new scheduling scheme. The model is placed in Figure 14, and the data are summarized in Tables 7 and 8.

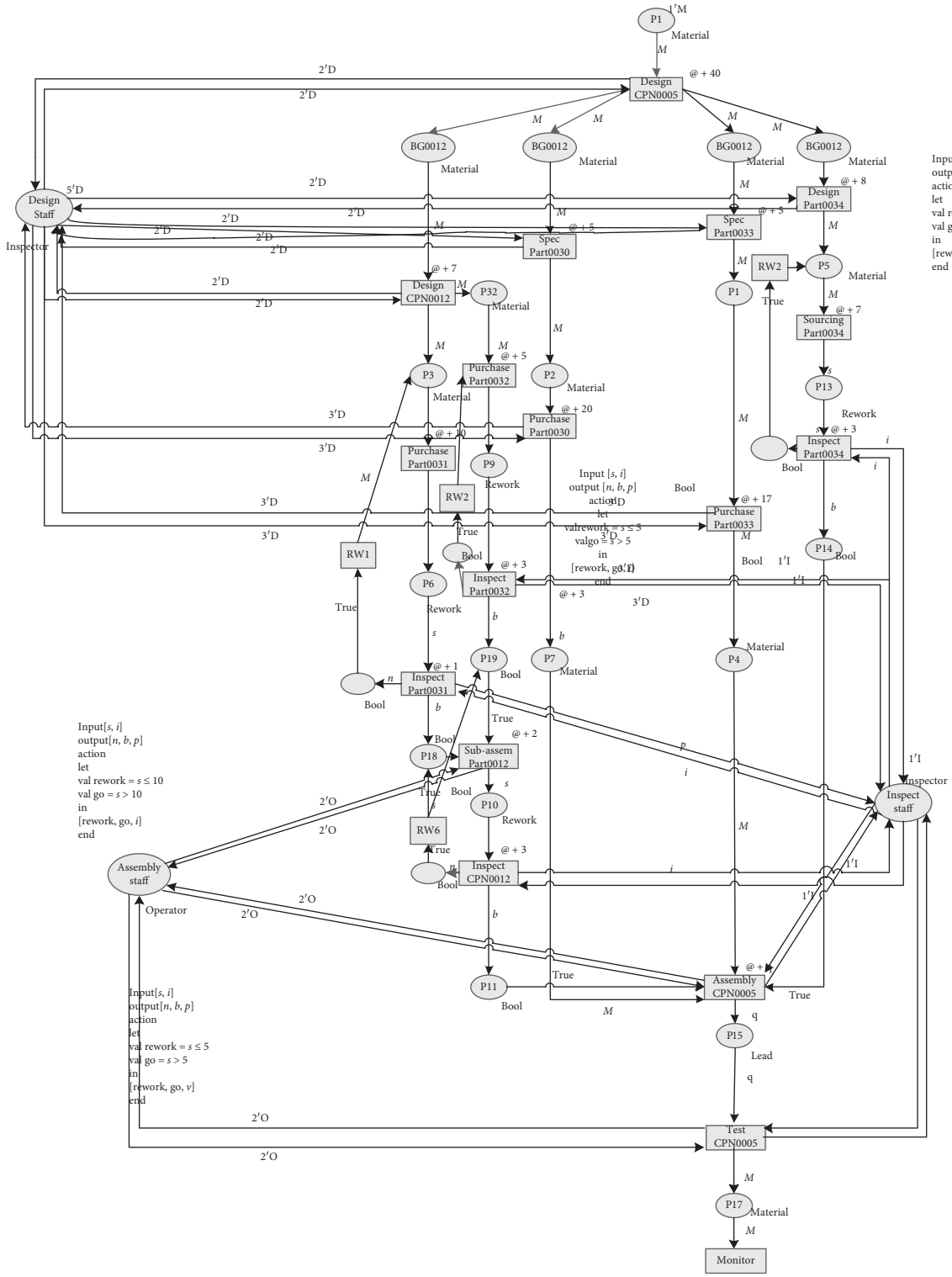


Figure 14: Optimized CPN model proposed.

Table 7: ESD comparison.

Task	AESD	TESD	LSD	DELAY	FAIL	WINDOW	ALSD	WDELAY
OTLT	109	100	100	9	9	0	202	102
NTLT	79	76	100	3	−21	24	131	31

Table 8: Reliability comparison.

Task	AVG	LSD	FAIL	STD	6 sigma
OTLT	118.06	100	18.06	11.06	51.24
NTLT	86.7	100	−13.3	6.67	6.71

As shown in Table 7, the NTLT performs much better in every aspect. The actual earliest starting date improves by 38%. The DELAY value for NTLT is smaller than OTLT, which means that the process runs more efficiently due to the provision of sufficient resources. The FAIL value for NTLT is now negative and, thus, it is now possible to complete the task on time. NTLT also has a window, which means that it is not tightly scheduled; hence, there is spare time to cope with emergencies such as rework. The WDELAY for NTLT is substantially shorter than OTLT, and therefore, the worst possible conditions for the new workflow will perform much better than the old one.

As shown in Table 8, the average lead time is significantly reduced for the new workflow. The FAIL value now becomes negative and the process is, hence, more likely to succeed. There is also a slight reduction in standard deviation for the new workflow; thus, the new process does a better job in risk control. The 6 sigma value drops from over 50 to approximately 6. This is a demonstration of how reliable this process becomes comparing to the old workflow. The new workflow lies in the plus and minus two sigma range and there is, therefore, a 95% successful rate for the entire process.

7. Conclusion

This paper has presented an approach for analyzing and improving the workflow for the Phobos-Grunt SOPSYS delivery unit manufacturing process via a colored Petri net (CPN).

First, the gap of the current study in workflow optimization and CPN implementation was identified. There is no implication of CPN in the literature available for a space project workflow optimization. This paper filled this gap by presenting a comprehensive approach to analyze, optimize, and validate an existing time-sensitive, single unit, space instrument manufacturing workflow via a CPN. Second, this paper constructed a CPN model for the specific case. Some extended CPN theories were introduced to support the construction model. We have proposed a framework and method of analysis for the construction of the model. For example, a CPN structure for rework logic is developed to simulate the real case. Third, this research generated an improvement proposal according to the simulation data: the ECRS method and theory of constraints theories are used for generating the proposal.

The first practical implication of this work is the Phobos-Grunt SOPSYS delivery unit manufacturing process itself. The ultimate goal of this study is to reduce the manufacturing process lead time in the space project. The results of the study have demonstrated its achievement. This research helps avoid the potential failure in the SOPSYS delivery unit manufacturing process and improves the reliability of it with a well-developed improvement strategy. The second implication of this study is its extended value for all the other normal time-sensitive single-unit manufacturing processes. Although the case modeling focuses on the Phobos-Grunt SOPSYS delivery unit manufacturing process, the logic of the process is similar to most other time-sensitive single-unit manufacturing processes. The same process and techniques can be used in other manufacturing systems for workflow optimization purposes in various space missions to the moon, Mars, and other planets.

Due to the relatively specific research background of this article, the data related to resources and time were obtained from the real SOPSYS unit manufacturing system. For this precise and customized instrument manufacturing process, accurate resource positioning must be achieved. Second, resource management systems often need to be manually modified for emergencies and resiliences. The resource management system could be built for a long-term, stable, and resource-intensive manufacturing system: the framework and techniques for building such a system are being further refined and developed using this case study's results as a basis. Overall, our research has provided some theory and experiences for workflow optimization in various space missions in future.

Data Availability

The data used in the article have been given in Table 3 of Section 4. The background information support this study can be found in "Polyu.edu.hk (2015) life on mars? award-wining polyu device could dig up the answer" (https://www.polyu.edu.hk/openingminds/en/story.php?sid=3).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

All authors has been contributed equally to this work.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under grant nos. 71772033, 71831003, and 71801031, in part by the Scientific and Research Funds of the Education Department of Liaoning Province of China under grant no. LN2019Q14, in part by the Natural Science Foundation of Liaoning Province of China (joint open fund for key scientific and technological innovation bases) under grant no. 2020-KF-11-11, and in part by the Department of Industrial and Systems Engineering of the Hong Kong Polytechnic University under grant no. H-ZG3K

References

[1] T. M. Koulopoulos, *The Workflow Imperative*, Van Nostrand Reinhold, New York, NY, USA, 1995, https://www.researchgate.net/publication/229124468_The_Workflow_Imperative.

[2] P. Lawrence, *Workflow Management Coalition, Workflow Handbook*, John Wiley & Sons, Hoboken, NJ, USA, 1997, https://dl.acm.org/citation.cfm?id=272945.

[3] H. Hong-Jun and Y. Jue, "An airport pavement traffic simulation based on CPN," *International Journal of Engineering Business Management*, vol. 6, 2014.

[4] W. Tan, W. Xu, F. Yang et al., "A framework for service enterprise workflow simulation with multi-agents cooperation," *Enterprise Information Systems*, vol. 7, pp. 523–542, 2013.

[5] Y. Q. Lv, C. K. M. Lee, Z. Wu et al., "Priority-based distributed manufacturing process modeling via hierarchical timed color petri net," *IEEE Transactions on Industrial Informatics*, vol. 9, pp. 1836–1846, 2013.

[6] J. W. Wang, W. H. Ip, R. R. Muddada et al., "On Petri net implementation of proactive resilient holistic supply chain networks," *The International Journal of Advanced Manufacturing Technology*, vol. 69, pp. 427–437, 2013.

[7] R. Y. K. Fung, A. Y. M. Au, and A. W. H. Ip, "Petri Net-based workflow management systems for in-process control in a plastic processing plant," *Journal of Materials Processing Technology*, vol. 139, pp. 302–309, 2003.

[8] J. L. Peterson, *Petri Net Theory and the Modeling of Systems*, Prentice-Hall, Upper Saddle River, NJ, USA, 1981.

[9] K. Jensen, "Coloured petri nets and the invariant-method," *Theoretical Computer Science*, vol. 14, pp. 317–336, 1981.

[10] K. Yung, C. W. Lam, S. M. Ko et al., "The Phobos-Grunt microgravity soil preparation system," *Acta Astronaut*, vol. 141, pp. 22–29, 2017.

[11] K. V. Hee, O. Oanea, R. Post et al., "Yasper: a tool for workflow modeling and analysis," in *Proceeding of the Sixth International Conference on Application of Concurrency to System Design (ACSD'06)*, pp. 279–282, Turku, Finland, 2006.

[12] J. Kurt, "A brief introduction to coloured petri nets," *Tools & Algorithms for Construction & Analysis of Systems*, Springer, Berlin, Heidelberg, 1997.

[13] R. Dewan, A. Seidmann, and Z. Walter, "Workflow optimization through task redesign in business information processes," in *Proceeding of the Hawaii International Conferenceon System Sciences*, p. 240, IEEE, Kohala Coast, HI, USA, 1998.

[14] R. Bartk, M. Jaska, L. Novk et al., "Workflow optimization with FlowOpt: on modelling, optimizing, visualizing, and analysing production workflows,"in *Proceeding of the International Conference on Technologies and Applications of Artificial Intelligence*, pp. 167–172, IEEE, Chung-Li, Taiwan, 2011.

[15] F. Crop, T. Lacornerie, X. Mirabel et al., "Workflow optimization for robotic stereotactic radiotherapy treatments: application of constant work in progress workflow," *Operations Research for Health Care*, vol. 6, pp. 18–22, 2015.

[16] Y. Du, Z. Yu, B. Yang et al., "Modeling and simulation of time and value throughputs of data-aware workflow processes," *Journal of Intelligent Manufacturing*, vol. 30, no. 6, pp. 2355–2373, 2019.

[17] B. Karlik, M. Uzam, M. Cinsdikici et al., "Neurovision-based logic control of an experimental manufacturing plant using neural net le-net5 and automation petri nets," *Journal of Intelligent Manufacturing*, vol. 16, no. 4-5, pp. 527–548, 2005.

[18] C. Kasemset, P. Pinmanee, and P. Umarin, "Application of ECRS and simulation techniques in bottleneck identification and improvement: a paper package factory industry 4.0 for SMEs-EU horizon 2020 RISE program view project application of ECRS and simulation techniques in bottleneck identification and improvement: a paper package factory," in *Proceedings of the Asia Pacific Industrial Engineering & Management Systems Conference*, Hong Kong, China, December 2018, https://www.researchgate.net/publication/276417414.

[19] M. À Piera, M. Narciso, A. Guasch et al., "Optimization of logistic and manufacturing systems through simulation: a colored petri net-based methodology," *Simulation*, vol. 80, pp. 121–129, 2004.

[20] N. Wightkin and U. Buy, "Formal modeling of sequential function charts with time Petri nets," *IEEE Transactions on Control Systems Technology*, vol. 20, pp. 678–692, 2014.

[21] W. M. Zuberek and W. Kubiak, "Timed petri nets in modeling and analysis of simple schedules for manufacturing cells," *Computers & Mathematics with Applications*, vol. 37, no. 11–12, pp. 191–206, 1999.

[22] E. Villani and J. C. Pascal, "A petri net-based object-oriented approach for the modelling of hybrid productive systems," *Nonlinear Analysis*, vol. 62, no. 8, pp. 1394–1418, 2005.

[23] M. Gao, "Modeling, service planning and service composition in knowledge-intensive collaborative workflows," *Ph.D.dissertation, School Management Science and Engineering*, Dongbei University of Finance and Economics, Dalian, China, 2013.

[24] D. R. Kiran, "Six sigma," in *Total Quality Management*, Elsevier, Amsterdam, Netherlands, pp. 347–361, 2016.

[25] D. Y. Mo, S. C. H. Ng, and D. Tai, "Revamping NetApp's service parts operations by process optimization," *INFORMS Journal on Applied Analytics*, vol. 49, no. 6, pp. 407–421, 2019.

[26] V. Rafiee and J. Faiz, "Robust design of an outer Rotor Permanent magnet motor through six-sigma methodology using Response surface surrogate model," *IEEE Transactions on Magnetics*, vol. 55, no. 10, pp. 1–10, 2019.

[27] L. G. Mkhaimer, M. Arafeh, and A. H. Sakhrieh, "Effective implementation of ISO 50001 energy management system: applying Lean six sigma approach," *International Journal of Business Management*, vol. 9, pp. 1–12, 2017.

[28] Y. N. Al Khamisi, M. Khurshid Khan, and J. E. Munive-Hernandez, "Assessing quality management system at a tertiary hospital in Oman using a hybrid knowledge-based system," *International Journal of Engineering Business Management*, vol. 10, pp. 1–13, 2018.

[29] F. A. A. Miranda, "Application of work sampling and ECRS (eliminate, combine, Re-lay out and simplify) principles of improvement at TO1 Assembly," in *Proceedings fo the 21stASEMEP National Technical Symposium*, Las Vegas, NV, USA, 2018, http://www.onsemi.com/site/pdf/SANYO-ApplicationofWorkSamplingandECRS-FINAL2.pdf.

[30] T. Kis, D. Kiritsis, P. Xirouchakis et al., "A petri net model for integrated process and job shop production planning," *Journal of Intelligent Manufacturing*, vol. 11, no. 2, pp. 191–207, 2000.

[31] W. M. Zuberek and W. Kubiak, "Timed petri nets in modeling and analysis of simple schedules for manufacturing cells," *Computers & Mathematics with Applications*, vol. 37, no. 11-12, pp. 191–206, 1999.

[32] X. You, L. Zhang, and J. Jiao, "Supply chain configuration modeling based on colored Petri-nets," in *Proceedings of the*

IEEE International Conference on Management of Innovation and Technology, pp. 921–925, IEEE, Kauia, HI, USA, 2006.

[33] M. Y. Marov, V. S. Avduevsky, E. L. Akim et al., "Phobos-Grunt: Russian sample return mission," *Advances in Space Research*, vol. 33, pp. 2276–2280, 2004.

[34] Phobos-Grunt's sad return," *Astronomy & Geophysics*, vol. 53, pp. 10-11, 2012.

[35] K. Jensen, L. M. Kristensen, and L. Wells, "Coloured Petri Nets and CPN Tools for modelling and validation of concurrent systems," *International Journal on Software Tools for Technology Transfer*, vol. 9, pp. 213–254, 2007.

[36] CPN Tools, 2018, http://cpntools.org/.

[37] Monitors–CPN Tools, 2018, http://cpntools.org/2018/01/12/monitors/.

[38] W. P. Oosterhuis and M. J. Severens, "Performance specifications and six sigma theory: clinical chemistry and industry compared," *Clinical Biochemistry*, vol. 57, pp. 12–17, 2018.

Hindawi
Discrete Dynamics in Nature and Society
Volume 2020, Article ID 9634308, 11 pages
https://doi.org/10.1155/2020/9634308

Research Article

Indicator Selection for Topic Popularity Definition Based on AHP and Deep Learning Models

Yuling Hong[1,2] **and Qishan Zhang**[1]

[1]*School of Economics and Management, Fuzhou University, Fuzhou 350108, China*
[2]*Computer Engineering College, Jimei University, Xiamen 361021, China*

Correspondence should be addressed to Qishan Zhang; zhangqs@fzu.edu.cn

Received 19 April 2020; Revised 24 June 2020; Accepted 6 July 2020; Published 24 August 2020

Guest Editor: Fuqiang Gu

Purpose. The purpose of this article is to predict the topic popularity on the social network accurately. Indicator selection model for a new definition of topic popularity with degree of grey incidence (DGI) is undertook based on an improved analytic hierarchy process (AHP). *Design/Methodology/Approach*. Through screening the importance of indicators by the deep learning methods such as recurrent neural networks (RNNs), long short-term memory (LSTM), and gated recurrent unit (GRU), a selection model of topic popularity indicators based on AHP is set up. *Findings*. The results show that when topic popularity is being built quantitatively based on the DGI method and different weights of topic indicators are obtained from the help of AHP, the average accuracy of topic popularity prediction can reach 97.66%. The training speed is higher and the prediction precision is higher. *Practical Implications*. The method proposed in the paper can be used to calculate the popularity of each hot topic and generate the ranking list of topics' *popularities*. Moreover, its future popularity can be predicted by deep learning methods. At the same time, a new application field of deep learning technology has been further discovered and verified. *Originality/Value*. This can lay a theoretical foundation for the formulation of topic popularity tendency prevention measures on the social network and provide an evaluation method which is consistent with the actual situation.

1. Introduction

On a social network platform, the user's behaviours such as posting or reposting or commenting on the content of a certain post may be related to one or more topics. The degree to which a topic is concerned by users is called topic popularity.

Recently, researches pointed out that the time sequence of topic popularity in the social network can be used to predict the development of topic tendency [1, 2]. Existing topic popularity predictions are mainly based on records of one indicator which is saved as historical data for linear regression and other methods. The indicator of the user's behaviours related to the topic popularity that changes over time can be regarded as time series. Research has shown that records of these indicators can be beneficial, provided that there are the numbers of reposts, comments, likes, and so on [3, 4]. Moreover, research found that topic popularity was affected by a combination of indicators [5, 6]. In recent years, some scholars have proposed the definition of topic popularity based on comprehensive multiindicators, but the influence of their weights on the popularity has not been studied [7],. and algorithms for topic popularity perdition are mainly based on the historical series data employing linear regression analysis methods. The rapid increase in data volume has led to the problems of inaccurate prediction results and poor system stability in traditional prediction systems for a long time [8]. Therefore, it is necessary to analyse historical topic popularity with many other indicators, especially the time series indicators, and to use nonlinear analysis methods to extract key indicators that affect topic popularity so as to predict the future development of it.

In order to make an appropriate and effective assessment, an indicator selection hierarchical structure of AHP for topic popularity prediction is applied. Therefore, in this paper, an indicator selection model is proposed with AHP for topic popularity definition. After that, this new definition is used in topic popularity prediction process which employs the deep learning algorithms for popularity series forecasting. In this way, the accuracy of forecasting results by three kinds of deep leaning models, which are popular and good at time series processing, is used to screen and weight the indicators.

The contributions of this paper are summarized as follows. In the second section, topic popularity is firstly defined with the concept of DGI. After that, the basic procedure of AHP is introduced for indicator selection for topic popularity and improved with the deep learning prediction models used mostly in time series problem. The third section explains the AHP-based popularity indicator extraction model proposed in this paper. The fourth section uses three kinds of most popular deep learning algorithms for topic popularity prediction and conducts experiments and analysis. The last section explains the conclusions and points out the future research directions of this article.

2. Preliminaries

Indicator analysis of topic popularity is used to evaluate the importance of the indicators that affect topic popularity.

2.1. Topic Popularity Indicator Selection. The core and goal of big data mining on social network is prediction. Establishing an indicator selection system is the basis for topic popularity prediction, so it is necessary to carry out topic popularity prediction based on monitored data and dynamically modify the prediction results with the addition of new data to achieve dynamic prediction of topic popularity. At present, the indicators that measure the popularity of social network topics are related to those whose values could intuitively rank the topic popularity by users' behaviours, which objectively reflects the tendency of social network topic popularity. Current research on topic popularity prediction is based on single or multiple indicators of topics. Wu and Huberman [9] defined the popularity of each post on the "Digg.com" website by counting the number of votes, through which the readers expressed their attitudes towards it. Their research found that the popularity followed a logarithmic normal distribution. In addition, Wu et al. [10] further studied the content decay law from the perspective of the number of users' comments and found that both the time interval between two consecutive comments on the same content and the frequency of comments on a topic obeyed a power law distribution. With the advancement of data mining technology on social network, Ratkiewicz et al. [11] studied the popularity of Wikipedia and counted the link penetration of nodes and found that it changed with time and it was more obeying the power law distribution. Lerman and Hogg [12] proposed to use random state transitions to represent users' registration, readings, likes, and behaviours of topic publishers or their friends on topics in social network. They assumed the number of votes on a story accumulates on Digg as its popularity and modelled the popularity of a post on the Digg as the independent variable and the number of likes received by the post at time t as the dependent variable. The rate equation of $N_{\text{vote}}(t)$ is as follows:

$$\frac{\mathrm{d}N_{\text{vote}}(t)}{\mathrm{d}t} = \text{rate}\left(\nu_f(t) + \nu_u(t) + \nu_{\text{friends}}(t)\right), \tag{1}$$

where rate means the probability at which a user seeing the story will vote on it, and ν_f, ν_u, and ν_{friends} are the rates at which users find the story via one of the front or upcoming pages and through the friends interface, respectively. These parameters are the empirical values trained from the training set. They found that when the users' behaviours (such as the number of likes) of the post is obvious enough, the topic post will be pushed to a page that can be seen by more users. The more likely its popularity is to increase; otherwise, the less popular the post will be and it may be gradually dissipated in social networks.

2.2. Basic Procedure of AHP. AHP is a multiobjective decision-making method that is practically applied in many engineering fields [13, 14]. It is a systematic analysis method combining qualitative analysis and quantitative analysis which was first proposed by Saaty [15]. Based on some defined criteria and procedure (see Figure 1), AHP technology mainly compares various indicators at the same level in complex problems in pairs and aims to determine the degree to which one alternative outperforms the other [16]. The basic procedure of AHP is as follows:

Step 1. Identify the problem and establish a hierarchical structure model.

By clarifying the scope of the problem, the specific requirements, the contained elements, and the relationship between each element, a complex multiple criteria decision-making (MCDM) problem is broken down into a hierarchy of interrelated decision elements according to the characteristics and general objectives of the problem. Generally, it is divided into three levels, which are the target layer at the top, the multiple criteria layer in the middle, and the alternatives layer which displays possible solutions or measures at the bottom.

Step 2. Build a judgment matrix.

Judgment matrix is the basis for relative importance calculation and hierarchical ordering. Based on an element C of the previous level as an evaluation criterion, the elements of each level are compared in pairwise to calculate their relative importance so as to determine the judgment matrix [16]. The elements to be compared and judged must have the same properties and be comparable. Based on n criteria, it requires $n(n-1)/2$ pairs of comparisons, where $N = [1, 2, \ldots, n]$ is the number of elements. Let $C = \{C_j \mid j = 1, 2, \&, n\}$ be the set of criteria and a $(n \times n)$ evaluation matrix A be

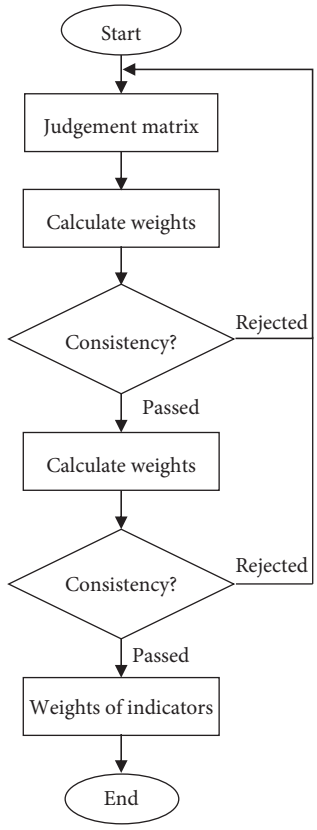


Figure 1: Flow chart of AHP.

the results of pairwise comparisons which is based on n criteria. The element a_{ij} in the evaluation matrix represents the relative importance of the element a_i to a_j according to the evaluation criterion C. The value of a_{ij} is determined after repeated researches on the data, expert opinions, and the experience of the evaluation subject. Each element a_{ij} $(i, j, = 1, 2, \ldots, n)$ is a quotient of weights of the criteria, as shown in

$$A = \left(a_{ij}\right)n \times n, \quad a_{ii} = 1,\ a_{ji} = \frac{1}{a_{ij}},\ a_{ij} \neq 0. \tag{2}$$

In AHP, multiple pairwise comparisons are based on a standardized comparison scale of nine levels as indicated in Table 1.

Step 3. Check the consistency.

In AHP, the indicator for judging matrix consistency is shown in the following equation:

$$\mathrm{CI} = \frac{\lambda_{\max} - n}{n - 1}, \tag{3}$$

where $\lambda_{\max}$ is the maximum eigenvalue of judgment matrix A and n is the order of judgment matrix A.

Step 4. Calculate the weight of each evaluation and make the decision.

Table 1: Scores of importance scale and its definition [12].

Scores of importance	Definition of importance
1	Equally important
3–1/3	Moderately more/less important
5–1/5	Strongly more/less important
7–1/7	Very strongly more/less important
9–1/9	Extremely more/less important
2–1/2, 4–1/4, 6–1/6, and 8–1/8	Intermediate values

The consistency ratio (CR) can be used to infer whether the assessment is sufficiently consistent, calculated by dividing the consistency index (CI) by the random consistency index (RI), as shown in the following equation:

$$\mathrm{CR} = \frac{\mathrm{CI}}{\mathrm{RI}}, \tag{4}$$

where RI is computed from Table 2 based on total number of random samples [17].

The greater the value of the CI, the greater the degree to which the judgment matrix deviates from full consistency. For verifying the calculated weights, Saaty [18] suggested that the value of CR should be less than 0.10 (maximum threshold). In the current work, CR values greater than 0.10 are rejected and a new pairwise comparison judgment is required. Finally, decision is made based on the normalized values mentioned above.

Research has also shown that the timing of topics was very important for their popularity, especially the social types of users [19]. By analysing and extracting user features and text features of participating topics, the method of machine learning can be used to forecast the influence of Weibo topics [20].

Recent advances in deep learning, especially the memory function of RNN on previous output sequences [21], provides some useful perspectives on how to solve time series problems. Zhu et al. [22] proposed an RNN opinion dynamics model for the prediction of each user's posting behaviours on Twitter and used an attention mechanism to predict user-level positions and merged the context of the neighbors' topics into a signal of interest. Although their approach can also be used to predict the topics from user-level, the RNN does not take the time sequence correlation from topic-level into consideration. According to the basic principles of deep learning methods, what is needed for solving the training problem is a reasonable input-to-output model and suitable amount of data for learning [23].

2.3. RNN Prediction Method. RNN is used for mining data for deep representation of time series information and semantic information [24]. It is commonly used in speech recognition, language modeling, machine translation, and time series analysis. The difference between RNN and ordinary fully connected neural networks is that the nodes

Table 2: Random consistency index (RI) for pairwise comparison.

N	3	4	5	6	7	8	9	10	11	12
RI	0.58	0.89	1.12	1.24	1.32	1.41	1.45	1.49	1.51	1.54

between the hidden layers of RNN are connected. The input of the hidden layer includes the output of the input layer and the output of the hidden layer at the previous moment. A recurrent neural network is essentially a type of neural network with an internal loop. The way it processes sequences is to iterate through all the elements in the sequence and retain some state information related to the viewed content. The structure of RNN employed in this study has two inputs x_1 and x_2, two hidden layers h_1 and h_2, an output layer Y, and the weights u, v, w and biases b_1 and b_2 on them (see Figure 2).

Suppose that h_1 and h_2 are used to represent the ith hidden layer in the hidden layers, v and u are used to indicate the weights for input nodes x_1 and x_2, and b_i is used to represent the offset term coefficient corresponding to the ith output node, then the output of the prediction model is shown in (7) below:

$$h_1 = v \times h_0 + u \times x_1 + b_1, \tag{5}$$

$$h_2 = v \times h_1 + u \times x_2 + b_1, \tag{6}$$

$$\hat{y} = w \times h_2 + b_2, \tag{7}$$

$$F(W, V, U, B) = \frac{1}{2}(\hat{y} - y)^2. \tag{8}$$

2.4. LSTM Prediction Method. LSTM is a type of RNN. Unlike traditional feedforward neural networks, RNN is a time series-based model that can establish the time correlation between previous information and the current environment. LSTM, proposed by Hochreiter and Schmidhuber [25] in 1997, mainly solves the problem of data classification and is applied to many aspects such as natural language processing, image subtitles, and speech recognition [26]. Since it can perfectly simulate the problem of multiple input variables, it can also be used for time series prediction [21].

LSTM adds a memory unit dedicated to storing historical information. The schematic diagram of the LSTM memory unit is shown in Figure 3, where $h^{(t)}$ is the hidden state vector for the current moment t while $\tilde{h}^{(t)}$ is the candidate state obtained with a hyperbolic tangent. $x^{(t)}$ is the current input vector. Historical information is updated through the control of three gates: input gate, forget gate, and output gate.

In this paper, it means popularity indicator or output vector of previous layer. "σ" and "tan$h$" represent separately for sigmoid function and hyperbolic tangent activation function. "$\circ$" stands for the element-wise multiplication. $i^{(t)}$, $f^{(t)}$, $o^{(t)}$, $\tilde{c}^{(t)}$, and $c^{(t)}$ are the input gate, forget gate, output gate, new memory cell, and final memory cell used in the LSTM model. Topic popularity time sequence is used as an input of the model. Its model is defined by the following equations (9)–(15) (for simplicity, the layer index l is omitted):

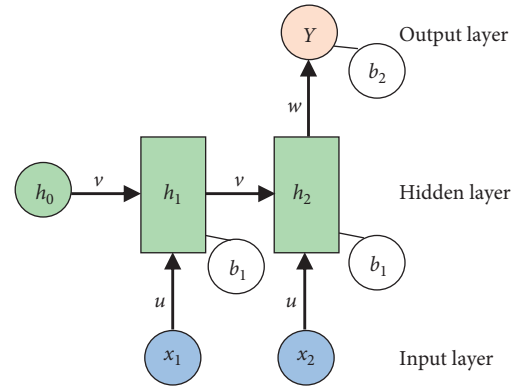


Figure 2: Structure diagram of RNN with two hidden layers.

$$i^{(t)} = \sigma\left(V^{(i)} \times h^{(t-1)} + U^{(i)} \times x^{(t)}\right), \tag{9}$$

$$f^{(t)} = \sigma\left(V^{(f)} \times h^{(t-1)} + U^{(f)} \times x^{(t)}\right), \tag{10}$$

$$o^{(t)} = \sigma\left(V^{(o)} \times h^{(t-1)} + U^{(o)} \times x^{(t)}\right), \tag{11}$$

$$\tilde{c}^{(t)} = tanh\left(V^{(c)} \times h^{(t-1)} + U^{(c)} \times x^{(t)}\right), \tag{12}$$

$$c^{(t)} = f^{(t)} \circ c^{(t-1)} + i^{(t)} \circ \tilde{c}^{(t)}, \tag{13}$$

$$h^{(t)} = o^{(t)} \circ \tanh\left(c^{(t)}\right) + i^{(t)} \circ \tilde{c}^{(t)}, \tag{14}$$

$$\hat{y} = w \times h^{(t)} + b. \tag{15}$$

2.5. GRU Prediction Method. The structure diagram of a GRU node is shown in Figure 4. Its model is defined by the following equations (16)–(20) (for simplicity, the layer index l is also omitted) [21]. $r^{(t)}$, $\tilde{h}^{(t)}$, and $z^{(t)}$ stand for reset gate, the candidate state obtained with a hyperbolic tangent, and update gate. The other symbols are basically the same as those used in the former two models.

$$r^{(t)} = \sigma\left(V^{(r)} \times h^{(t-1)} + U^{(r)} \times x^{(t)}\right), \tag{16}$$

$$z^{(t)} = \sigma\left(V^{(z)} \times h^{(t-1)} + U^{(z)} \times x^{(t)}\right), \tag{17}$$

$$\tilde{h}^{(t)} = \phi\left[V^{(h)} \times \left(r^{(t)} \circ h^{(t-1)}\right) + U^{(h)} \times x^{(t)}\right], \tag{18}$$

$$h^{(t)} = \left(1 - z^{(t)}\right) \times \tilde{h}^{(t)} + z^{(t)} \times h^{(t-1)}, \tag{19}$$

$$\hat{y} = w \times h^{(t)} + b. \tag{20}$$

Given that indicator sets are formed by experts, decision makers may need to analyse large amounts of data and consider many indicators [7].

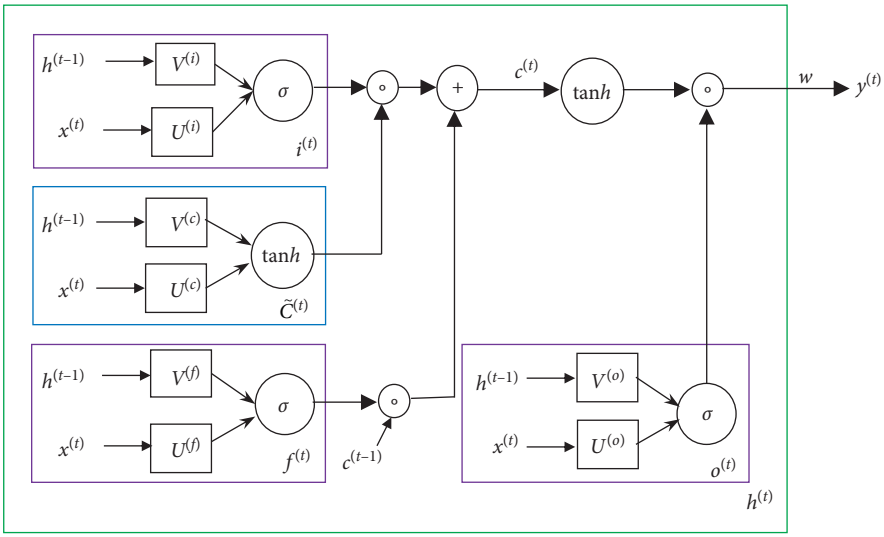


Figure 3: Structure diagram of a LSTM node.

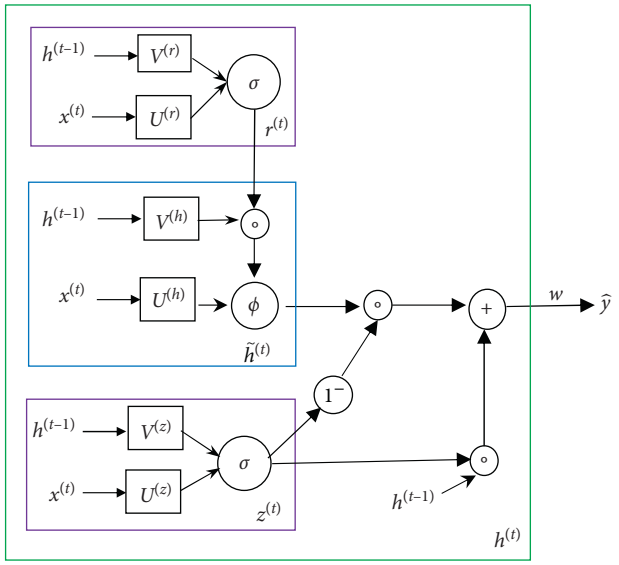


Figure 4: Structure diagram of a GRU node.

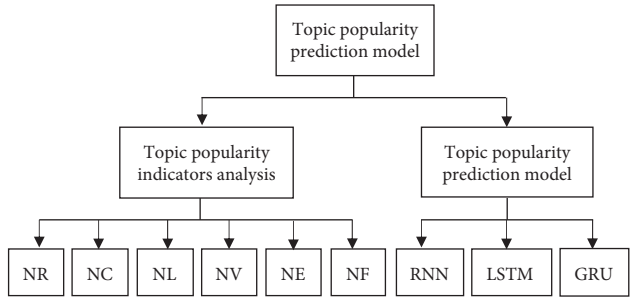


Figure 5: Topic popularity prediction model.

3. Topic Popularity Prediction Model

In this paper, the topic popularity prediction model considers two aspects, including topic popularity indicators analysis and topic popularity prediction (see Figure 5). Combining the prominent characteristics of the popularity of social network topics and the fact that exists in many social networks at present [4, 6, 27], the indicators which influence the topic popularity can be roughly divided into six indicators as follows. There is no overlap or intersection between levels, and the measurement scheme of each evaluation indicator is easy to implement and has strong operability [6].

NR (number of retweets): users on the social network can forward a topic to their personal homepages and comment on it. It refers to the number of times a topic has been retweeted during the period from the moment it is published to the moment it is measured.

NC (number of comments): it refers to the number of readers' opinion or reaction to the content of the topic, which is expressed after it is published.

NL (number of likes): it refers to the number of times a topic is clicked as "Likes" by a network user from the moment it is published to the moment it is measured.

NV (number of views): it refers to the number of times that a topic is read by users from the moment it is published to the moment it is measured.

NE (number of related entities): it refers to the number of the tags of the entities or people's names mentioned from the moment it is published to the moment it is measured in a topic.

NF (number of favourites): it refers to the number of times a topic is favorited from the moment it is published to the moment it is measured by a network user.

3.1. Definition of Topic Popularity in Social Network. In social network, the concept of topic heat refers to the comprehensive value of indicators of network user's behaviour in the dissemination of topic's information. The calculation method for topic heat at the moment t is shown as

$$H_t^* = w_1 a_t + w_2 b_t + w_3 c_t + w_4 d_t + w_5 e_t + w_6 f_t, \quad (21)$$

where the w_i, $i = 1, 2, \ldots, 6$, stands for different weights for each indicator. Table 3 shows data of topic heat based on users' behaviour indicators.

In order to make the data comparable and reliable and to eliminate the difference of the variation between indicators, we adopt the range normalization procedure on the original data of topic's indicators. In order to control the topic heat to [0, 100] after this normalization, the value is multiplied by 100 to obtain the topic heat H_t, and it is calculated from equation (22) as follows:

$$H_t = \frac{H_t^* - \min(H_t^*)}{\max(H_t^*) - \min(H_t^*)} \times 100, \quad (22)$$

where $\max(H_t^*)$ and $\min(H_t^*)$ respectively represent the maximum and minimum values of topic heat in the original data.

Let H be the time series vector of topic heat, then

$$H = (H_1, H_2, \ldots, H_t). \quad (23)$$

Since the popularity of social network topics presents a discrete state over time, first, according to the popularity vector obtained by equation (23), the acceleration value of the topic heat at each moment is calculated as

Table 3: Data of topic heat based on users' behaviour indicators.

Moment (T)	A (NR)	B (NC)	C (NL)	D (NV)	E (NE)	F (NF)
1	a_1	b_1	c_1	d_1	e_1	f_1
2	a_2	b_2	c_2	d_2	e_2	f_2
...	...	...	...	...	...	...
t	a_t	b_t	c_t	d_t	e_t	f_t

$$\Delta H_i = H_{i+1} - H_i, \quad i = 1, 2, \ldots, t-1. \quad (24)$$

Considering the topic heat acceleration as a time series, let ΔH be the heat acceleration vector,

$$\Delta H = (\Delta H_1, \Delta H_2, \ldots, \Delta H_{t-1}). \quad (25)$$

Suppose there are a total of n topics in the period of [0, T], then the heat tendency of topic i ($i = 1, 2, \ldots, n$) at time t is composed by the current H_i and the heat acceleration ΔH_i, as shown in Figure 6:

$$\overline{H}_i(t) = \frac{1}{1+\alpha} \cdot H_i(t) + \frac{\alpha}{1+\alpha} \cdot \Delta H_i(t-1). \quad (26)$$

Here, $\alpha = 0.4$ is used [28]. Among them, the definition of $\overline{H}_i'$ is calculated by the degree of grey incidence (DGI) [29] according to

$$\overline{H}_i'(t) = \frac{\min_i \min_t \left|\max_{1\le i\le n} \overline{H}_i(t) - \overline{H}_i(t)\right| + \rho \max_i \max_t \left|\max_{1\le i\le n} \overline{H}_i(t) - \overline{H}_i(t)\right|}{\left|\max_{1\le i\le n} \overline{H}_i(t) - \overline{H}_i(t)\right| + \rho \max_i \max_k \left|\max_{1\le i\le n} \overline{H}_i(t) - \overline{H}_i(t)\right|}, \quad (27)$$

in which $\rho \in [0.1, 0.5)$. Thus, at the moment t, the popularity of the ith topic on the social network is shown as follows:

$$\text{popularity}_i = \sum_{k=1}^{n} w_i(k) * \overline{H}_i'(k), \quad (28)$$

where the w_i stands for different weights for each indicator.

3.2. AHP for Topic Popularity Indicator Selection

(1) A hierarchical structure model is established based on the definition of topic popularity. As shown in Figure 6, the indicator selection model for topic popularity prediction is divided into three layers, of which "*Goal*" layer represents the evaluation target layer, that is, the indicator selection for topic popularity prediction; "*Indicators*" layer represents the indicator selected from the literatures, which includes "NR," "NC," "NL," "NV," "NE," and "NF"; the prediction results of the deep learning prediction models on the topic indicators are employed as the screening principle for the topic popularity indicators. Deep learning models train and predict the popularity on each indicator and select the indicators with the highest accuracies among the prediction outputs.

(2) The data of the judgment matrix comes from the questionnaires of 10 experts and professors engaged in big data mining using deep learning algorithms in social networks for several years. When they compare the indicators and criteria and strategies, they use their perception and experiments, alongside with the objective facts at their disposal. For example, 2.08 in Table 4 indicates that the average score of the experts for the importance of $i1$ to $i2$ is 2.08. Judgment matrixes of the strategic layer relative to each criterion are shown in Tables 5–10.

(3) Calculate the parameters according to formulas (3) and (4) (see the textbook [30] for the specific process).

Table 11 shows the results obtained from AHP on topic attributes, where the ranking is "NR" > "NC" > "NF" > "NE" > "NV" > "NL."

Table 12 presents the results obtained from AHP on prediction models with indicator NR, where the ranking is "GRU" > "LSTM" > "RNN."

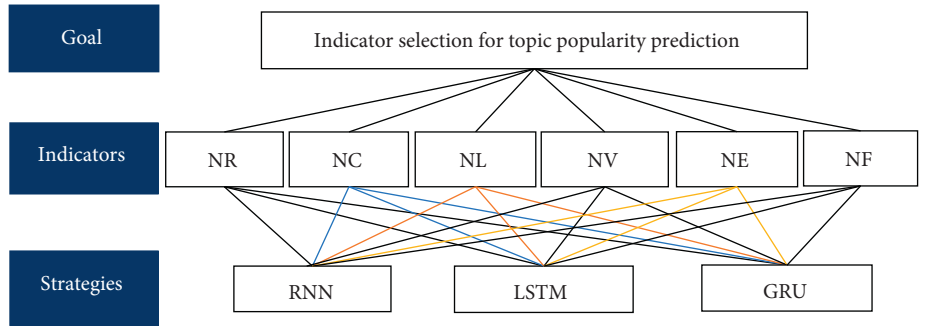


Figure 6: AHP-based indicator selection model for topic popularity prediction.

Table 4: Pairwise comparison judgment matrix of the indicator layer.

G	I_1	I_2	I_3	I_4	I_5	I_6
I_1	1	2.08	3.75	4.85	1.25	1.58
I_2	0.48	1	3.75	4.37	1.02	1.9
I_3	0.27	0.27	1	0.49	0.66	0.98
I_4	0.21	0.23	2.04	1	1.11	0.43
I_5	0.80	0.98	1.79	0.67	1	0.33
I_6	0.63	0.53	1.02	2.33	1.89	1

Table 5: Judgment matrix of the strategic layer relative to I_1.

I_1	S_1	S_2	S_3
S_1	1	0.75	0.55
S_2	1.33	1	0.86
S_3	1.82	1.16	1

Table 6: Judgment matrix of the strategic layer relative to I_2.

I_2	S_1	S_2	S_3
S_1	1	0.57	0.40
S_2	1.76	1	0.91
S_3	2.51	1.10	1

Table 7: Judgment matrix of the strategic layer relative to I_3.

I_3	S_1	S_2	S_3
S_1	1	0.28	0.52
S_2	3.56	1	0.97
S_3	1.93	1.03	1

Table 8: Judgment matrix of the strategic layer relative to I_4.

I_4	S_1	S_2	S_3
S_1	1	0.52	0.37
S_2	1.93	1	0.86
S_3	2.72	1.16	1

Table 9: Judgment matrix of the strategic layer relative to I_5.

I_5	S_1	S_2	S_3
S_1	1	0.79	0.39
S_2	1.27	1	0.91
S_3	2.59	1.10	1

Table 10: Judgment matrix of the strategic layer relative to I_6.

I_6	S_1	S_2	S_3
S_1	1	0.74	0.58
S_2	1.36	1	0.93
S_3	1.71	1.07	1

Table 11: Weightage of the indicators by AHP on topic attributes.

Indicators	Weightage of the indicators within each indicator by AHP						Weights
	1	2	3	4	5	6	
NR	1	2.08	3.75	4.85	1.25	1.58	0.309
NC	0.48	1	3.75	4.37	1.02	1.9	0.237
NL	0.27	0.27	1	0.49	0.66	0.98	0.080
NV	0.21	0.23	2.04	1	1.11	0.43	0.090
NE	0.80	0.98	1.79	0.67	1	0.33	0.124
NF	0.63	0.53	1.02	2.33	1.89	1	0.161
							CR = 0.0713

Table 12: Weightage of the indicator NR by AHP on prediction models.

Strategies	Weightage of the indicator NR by AHP			Weights
	1	2	3	
RNN	1	0.75	0.55	0.242
LSTM	1.33	1	0.86	0.340
GRU	1.82	1.16	1	0.417
				CR = 0.0024

Table 13: Weightage of the indicator NC by AHP on prediction models.

Strategies	Weightage of the indicator NC by AHP			Weights
	1	2	3	
RNN	1	0.57	0.40	0.192
LSTM	1.76	1	0.91	0.368
GRU	2.51	1.10	1	0.441
				CR = 0.0064

Table 14: Weightage of the indicator NL by AHP on prediction models.

Strategies	Weightage of the indicators NL by AHP			Weights
	1	2	3	
RNN	1	0.28	0.52	0.160
LSTM	3.56	1	0.97	0.458
GRU	1.93	1.03	1	0.382
				CR = 0.0399

Table 15: Weightage of the indicator NV by AHP on prediction models.

Strategies	Weightage of the indicators NV by AHP			Weights
	1	2	3	
RNN	1	0.52	0.37	0.179
LSTM	1.93	1	0.86	0.367
GRU	2.72	1.16	1	0.455
				CR = 0.0035

Table 16: Weightage of the indicator NE by AHP on prediction models.

Strategies	Weightage of the indicators NE by AHP			Weights
	1	2	3	
RNN	1	0.79	0.39	0.214
LSTM	1.27	1	0.91	0.334
GRU	2.59	1.10	1	0.452
				CR = 0.037

Table 17: Weightage of the indicator NF by AHP on prediction models.

Strategies	Weightage of the indicator NF by AHP			Weights
	1	2	3	
RNN	1	0.74	0.58	0.246
LSTM	1.36	1	0.93	0.354
GRU	1.71	1.07	1	0.400
				CR = 0.0026

Table 13 shows the results obtained from AHP on prediction models with indicator NC, where the ranking is "GRU" > "LSTM" > "RNN."

Table 14 presents the results obtained from AHP on prediction models with indicator NL, where the ranking is "LSTM" > "GRU" > "RNN."

In Table 15, the results are obtained from AHP on prediction models with indicator NV, where the ranking is "GRU" > "LSTM" > "RNN."

In Table 16, the results are obtained from AHP on prediction models with indicator NE, where the ranking is "GRU" > "LSTM" > "RNN."

In Table 17, the results are obtained from AHP on prediction models with indicator NF, where the ranking is "GRU" > "LSTM" > "RNN."

All these CRs are below 0.10. Based on these, we calculate the CI for each one of the tables above and proceed to measure the total consistency for the popularity prediction problem. For Table 11, it is equal to 0.0713; while for the other six tables, they amount to 0.0024 (Table 12), 0.0064 (Table 13), 0.0399 (Table 14), 0.0035 (Table 15), 0.0037 (Table 16), and 0.0026 (Table 17). The overall consistency index is 0.0399.

(4) The deep learning prediction algorithms are applied for indicator selection, choosing the indicators that have a greater impact on the topic popularity. The overall priority of alternative S_i equals the sum of the strategies of each priority multiplied by the individual priority of alternative S_i. For example, in Table 18, the overall priority for S_1 equals

$$\begin{aligned}&(0.309)(0.242) + (0.237)(0.192)\\ \% \quad &+ (0.080)(0.160) + (0.090)(0.179) + (0.124)(0.214)\\ &+ (0.161)(0.246) = 0.215.\end{aligned} \tag{29}$$

Similarly, all other priorities are calculated in this way. Therefore, we have priority for RNN = 21.5%, priority for LSTM = 36.0%, and priority for GRU = 42.5%.

The calculation process is as follows. Firstly, these three algorithms of deep learning are used to predict the topic popularity time series, respectively. Secondly, the prediction results are weighted to obtain the final prediction value. Thirdly, the index with the prediction accuracy of the previous 85% of the index set is picked up as the final indicators for popularity definition [13].

4. Experiment Results and Discussion

Using these deep learning models, we can predict the topic popularity on online social network at the next moment. In addition, the accuracies of the three models of deep learning for predicting topic popularity are compared, including predictions on the original attributes and the novel definition of popularity with DGI.

4.1. Data Collection. Topic popularity data sources used in this study include the numbers of "retweets," "comments," "likes," "views," "entities," and "favorites" mentioned above as in Table 4. These are used to study the accuracy, resolution, and characteristics in topic popularity prediction. In this research, we design an evaluation questionnaire to compare the importance of various indicators of topic heat on topic popularity and the comparison of the prediction accuracy of RNN, LSTM, and GRU in each indicator.

The dataset for this experiment was collected from some topics on the Hot Topics List of Weibo at 6 pm, 2020.04.21, to 6 pm, 2020.04.24, and crawler technology was used to collect the numbers of "retweet," "comments," "likes," "views," "entities," and "favorites" every 5 minutes. For this synthetic dataset, we use a generation process similar to that

Table 18: Calculations of overall priorities between alternative options.

Strategies	I_1(0.309)	I_2(0.237)	I_3(0.080)	I_4(0.090)	I_5(0.124)	I_6(0.161)	Priorities
S_1	0.242	0.192	0.160	0.179	0.214	0.246	0.215
S_2	0.340	0.368	0.458	0.367	0.334	0.354	0.360
S_3	0.417	0.441	0.382	0.455	0.452	0.400	0.425

Table 19: Experimental cases.

Case no.	Characteristic description
1	RNN
2	LSTM
3	GRU
4	AHP-based w_1* (RNN) + w_2* (LSTM) + w_3* (GRU) proposed in this paper

Table 20: Comparison of accuracy of experimental results.

I_i case	NR (%)	NC (%)	NL (%)	NV (%)	NE (%)	NF (%)	Popularity (%)	Average (%)
1	94.16	91.30	92.82	82.57	89.30	92.67	96.30	91.30
2	97.22	96.64	94.28	87.30	93.28	94.43	98.26	94.49
3	97.24	96.66	94.30	85.67	93.30	94.46	98.65	94.33
4 (this paper)	98.16	98.15	97.91	96.25	97.31	97.31	98.96	97.66

described in [7]. We use 12 different topics as the training dataset and collect a total of 62208 sequences. All data instances in the dataset are 20 moments long (10 moments for the input and 10 moments for the prediction).

4.2. Experimental Case. This study mainly includes four experimental cases. The first three experiments are all popular algorithms used for time series in deep learning, and the last one is a case of AHP-based indicator selection method using deep learning method proposed in this paper, as shown in Table 19.

4.3. Evaluation of Topic Popularity Prediction. In order to demonstrate the accuracy of the proposed prediction model, the actual value $h(x_i)$ and the predicted value $\widehat{y}_i$ can be compared. Equation (30) is the mean absolute error (MAE) used to examine the accuracy of deep learning models in this paper.

$$\mathrm{MAE}(X,h) = \frac{1}{m}\sum_{i=1}^{m}\left|h(x_i) - \widehat{y}_i\right|. \tag{30}$$

The performance of the three deep learning models is shown in Table 20, where the "Popularity" means the definition proposed in this paper and the bold text indicates the smallest prediction error among the error values in each row and column. Here, the accuracy is calculated as follows:

$$\text{accuracy} = 1 - \mathrm{MAE} * 100\%. \tag{31}$$

Prediction experiments on topic popularity are conducted based on three different deep learning models and the AHP-based model mentioned in this paper. Comparison of accuracy of experimental results is shown in Table 20, where case nos. 1 to 3 are basic RNN, LSTM, and GRU, which use the "linear" activation and 5000 iterations at each experiment on a topic. No. 4 is the indicator selection model proposed in this paper. According to the experimental results, the improved AHP indicator selection model proposed by this research has the highest prediction accuracy, with an average accuracy of 97.66% and the average accuracy of experiments 1 to 3 are, respectively, 91.30%, 94.49%, and 94.44%. In particular, experiment on the proposed popularity definition improves the prediction accuracy up to 98.96%, as shown in Table 20.

5. Conclusions and Future Work

In this paper, we have successfully applied the deep learning approaches to the challenging topic popularity prediction problem based on AHP which so far has not benefited from sophisticated deep learning techniques. We made a qualitative research on the influencing indicators of social network topic popularity in the context of big data and build an indicator selection model for improving comprehensive evaluation of AHP by deep leaning approaches, especially RNN, LSTM, and GRU. By defining the maximum correlation vector on the indicator, the definition of topic popularity is built quantitatively based on the DGI method, and different weights of topic indicators are obtained from the help of AHP. Some new indicators are used in this research, including the indicators of the number of views, the number of entities mentioned or associated with a topic, and the number of favorites. The average accuracy of experiments can reach 97.66%. The comparison of experiments shows that using the prediction accuracy of deep learning algorithms as the indicator screening principle for AHP, the training speed of the topic popularity prediction model is faster and its prediction accuracy is higher. The indicators mentioned in the previous literature can well predict the development tendency of topic popularity. According to the topic popularity defined by DGI and using the prediction

algorithm weight provided by AHP to give the prediction value, the prediction accuracy obtained is better. This method is superior to those prediction models established without screening and is closer to the actual measurement. The predicted value is closer to the actual value, which plays a certain reference role for the supervision and control of the topic popularity tendency.

For future work, we will explore different kinds of platforms from social networking sites, exploit community analysis algorithms, and parameter adjustments for topic popularity prediction.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the Educational Commission Project of Fujian Province, China, under Grant JAT170327 and in part by the Natural Science Foundation of Fujian Province, China, under Grant 2018J01791 and Grant 2018J01539. The authors sincerely thank the experts who participated in the questionnaire for their work and concern.

References

[1] J. Yao and X. Meng, "Research on hot topic of e-commerce microblog based on improved LDA algorithm," *Journal of Chongqing University of Technology (Natural Science)*, vol. 33, no. 12, pp. 184–188, 2019.

[2] Z. Y. Zheng, G. L. Jiang, X. J. Zhang, Z. F. Wang, and D. Li, "Research on the prediction algorithm for Sina popular microblog based on multi-features," *Journal of Chinese Computer Systems*, vol. 38, no. 3, pp. 494–498, 2017.

[3] S. Wu, "A discussion on theoretical base of internet opinion," in *Proceedings of the 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce*, pp. 4908–4911, Deng Long, China, August 2011.

[4] X. Cao, X. Zhuang, L. Liu, K. Fang, F. Duan, and S. Li, "Research on internet public opinion heat base on the response level of emergencies," *Chinese Journal of Management Science*, vol. 22, no. 3, pp. 82–89, 2014.

[5] A. Pal and S. Counts, "Identifying topical authorities in microblogs," in *Proceedings of the 4th ACM International Conference on Web Search and Data Mining (WSDM 2011)*, pp. 45–54, Hong Kong, China, February 2011.

[6] F. Sun, S. Cheng, X. Jin, and T. Ni, "A quantitative evaluation method for negative internet public opinion of government—taking Sina Weibo as an example," *Journal of Intelligence*, vol. 34, no. 8, pp. 137–141, 2015.

[7] Y. Lan, B. Liu, P. Zhang, Y. Xia, and H. Li, "The internet public opinion hot-degree dynamic prediction model oriented to big data," *Journal of Intelligence*, vol. 36, no. 6, pp. 105–110, 2017.

[8] H. Jin, "Design of microblog hot spot prediction system based on improved neural network algorithm," *Modern Electronics Technique*, vol. 41, no. 12, pp. 157–160, 2018.

[9] F. Wu and B. A. Huberman, "Novelty and collective attention," *Proceedings of the National Academy of Sciences*, vol. 104, no. 45, pp. 17599–17601, 2007.

[10] Y. Wu, C. Zhou, M. Chen, J. Xiao, and J. Kurths, "Human comment dynamics in on-line social systems," *Physica A: Statistical Mechanics and Its Applications*, vol. 389, no. 24, pp. 5832–5837, 2010.

[11] J. Ratkiewicz, F. Menczer, S. Fortunato, A. Flammini, and A. Vespignani, "Traffic in social media II: modeling bursty popularity," in *Proceedings of the 2nd IEEE International Conference on Social Computing (PASSAT 2010)*, pp. 393–400, IEEE, August 2010.

[12] K. Lerman and T. Hogg, "Using a model of social dynamics to predict popularity of news," in *Proceedings of the 19th International Conference on World Wide Web (WWW '10)*, pp. 621–630, Raleigh, NC, USA, April 2010.

[13] X. Zhou, K. Guo, G. Bai, and D. Song, "Improved AHP combined with BP neural prediction of coal and gas outburst," *Computing Techniques for Geophysical and Geochemical Exploration*, vol. 41, no. 1, pp. 121–127, 2019.

[14] S. Luthra, S. K. Mangla, L. Xu, and A. Diabat, "Using AHP to evaluate barriers in adopting sustainable consumption and production initiatives in a supply chain," *International Journal of Production Economics*, vol. 181, no. 11, pp. 342–349, 2016.

[15] R. W. Saaty, "The analytic hierarchy process-what it is and how it is used," *Mathematical Modelling*, vol. 9, no. 3–5, pp. 161–176, 1987.

[16] F. D. Felice, A. Petrillo, and M. Tricarico, "Using analytic hierarchy process to evaluate human performance," in *Proceedings of 12th International Conference on Modeling and Applied Simulation*, pp. 75–84, Athens, Greece, September 2013.

[17] E. Mastrocinque, F. J. Ramírez, A. Honrubia-Escribano, and D. T. Pham, "An AHP-based multi-criteria model for sustainable supply chain development in the renewable energy sector," *Expert Systems with Applications*, vol. 150, Article ID 113321, 2020.

[18] T. L. Saaty, "Fundamentals of decision making and priority theory with the analytic hierarchy process," *Analytic Hierarchy Process Series, RWS*, RWS Publications, Pittsburgh, PA, USA, 2000.

[19] Y. Ke, "Analysis and effect prediction of the influence factors of Sina micro-blog information dissemination," *Journal of Modern Information*, vol. 36, no. 3, pp. 22–26, 2016.

[20] J. X. Cao, J. Wu, J. Shi et al., "Sina microblog information diffusion analysis and prediction," *Chinese Journal of Computers*, vol. 37, no. 4, pp. 779–790, 2014.

[21] K. Cho, B. Van Merriënboer, C. Gulcehre et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*, pp. 1724–1734, Doha, Qatar, October 2014.

[22] L. Zhu, Y. He, and D. Zhou, "Neural opinion dynamics model for the prediction of user-level stance dynamics," *Information Processing & Management*, vol. 57, no. 2, p. 102031, 2020.

[23] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," *Advances in Neural Information Processing Systems*, vol. 4, pp. 3104–3112, 2014.

[24] M. Hou, D. Pi, and B. Li, "Similarity-based deep learning approach for remaining useful life prediction," *Measurement*, vol. 159, p. 107788, 2020.
[25] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
[26] M. Li, F. Lu, H. Zhang, and J. Chen, "Predicting future locations of moving objects with deep fuzzy-LSTM networks," *Transportmetrica A: Transport Science*, vol. 16, no. 1, pp. 119–136, 2020.
[27] Y. Zhang, J. Qi, B. Fang, and Y. Li, "Research on the index system of public opinion on internet for abnormal emergence," *Journal of Intelligence*, vol. 29, no. 11, pp. 71–75, 2010.
[28] T. Brants, A. Popat, F. Och, and J. Dean, "Large language models in machine translation," in *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pp. 858–867, Prague, Czech Republic, June 2007.
[29] A. S. Khuman, Y. Yang, and S. Liu, "Grey relational analysis and natural language processing to: grey language processing," *Journal of Grey System*, vol. 28, no. 1, pp. 88–97, 2016.
[30] P. G. Prastacos, "Multi-criteria decision making," *Managerial Decision Making: Theory and Practice*, pp. 124–126, Tsinghua University Press, Beijing, China, 2009.

Hindawi
Discrete Dynamics in Nature and Society
Volume 2020, Article ID 8105145, 13 pages
https://doi.org/10.1155/2020/8105145

Research Article

Fuzzy Theory-Based Data Placement for Scientific Workflows in Hybrid Cloud Environments

Zheyi Chen,[1] Xu Zhao,[2] and Bing Lin[3]

[1]*College of Engineering Mathematics and Physical Sciences, University of Exeter, Exeter EX4 4QF, UK*
[2]*Fujian Provincial Key Laboratory of Network Computing and Intelligent Information Processing, Fuzhou 350118, China*
[3]*College of Physics and Energy, Fujian Normal University, Fuzhou 350117, China*

Correspondence should be addressed to Bing Lin; wheellx@163.com

Received 11 June 2020; Accepted 28 July 2020; Published 17 August 2020

Guest Editor: Chi-Hua Chen

In hybrid cloud environments, reasonable data placement strategies are critical to the efficient execution of scientific workflows. Due to various loads, bandwidth fluctuations, and network congestions between different data centers as well as the dynamics of hybrid cloud environments, the data transmission time is uncertain. Thus, it poses huge challenges to the efficient data placement for scientific workflows. However, most of the traditional solutions for data placement focus on deterministic cloud environments, which lead to the excessive data transmission time of scientific workflows. To address this problem, we propose an adaptive discrete particle swarm optimization algorithm based on the fuzzy theory and genetic algorithm operators (DPSO-FGA) to minimize the fuzzy data transmission time of scientific workflows. The DPSO-FGA can rationally place the scientific workflow data while meeting the requirements of data privacy and the capacity limitations of data centers. Simulation results show that the DPSO-FGA can effectively reduce the fuzzy data transmission time of scientific workflows in hybrid cloud environments.

1. Introduction

With the widespread applications of Big Data technologies, the amount of data generated by modern network environments is greatly increasing. Therefore, traditional distributed computing modes such as grid computing may not meet the requirements of massive data processing. In recent years, cloud computing has emerged as a research hotspot [1–5], where hybrid cloud environments show the advantages of high sharing, high availability, and customization. Specifically, the hybrid cloud environments are composed of the data centers distributed in different geographical locations, including multiple private and public data centers [6]. On the one hand, the public cloud is good at providing high reliability and large capacity with the resource-sharing feature. On the other hand, the private cloud is adept at offering high flexibility and security, which guarantees data privacy during the work process.

Due to the complexity of the work process and increasing data volumes, scientific research studies with strict work steps cannot be managed manually. To address this problem, the workflow technology was proposed [7], where scientific workflows [8] can be used to manage, monitor, and execute these scientific processes. However, the amount of data involved in scientific workflows is commonly huge, which may need to be stored in the data centers in different geographical locations and transferred across data centers during the operation of scientific workflows. Therefore, it has become a research hotspot to effectively execute the data placement for scientific workflows in hybrid cloud environments under limited bandwidth conditions with the goal of reducing data transmission time [9–12].

Some researchers have contributed to addressing the problem of data placement for scientific workflows. Yuan et al. [13] proposed a k-means clustering algorithm-based data placement method, which utilized the dependency

among data and considered the load balancing in data centers. Reddy et al. [14] designed an entropy-based data placement strategy for enhancing the map-reduce performance in Hadoop clusters, where the k-means clustering algorithm was used to group different datasets. However, this method was not suitable for hybrid cloud environments with different capacities of data centers. Li et al. [15] designed a data placement solution for hybrid data centers, which can reduce the data transmission time. Also for hybrid data centers, the data placement strategy proposed in [16] reduced the data transmission volume and the data transmission times across data centers. However, they did not consider some essential factors in data placement, such as differences in data centers (e.g., capacities and bandwidth) and bandwidth fluctuations. Zheng et al. [17] and Cui et al. [18] developed a data placement scheme based on the GA, which may easily fall into the local optimal solution during operation. As for optimization objectives in data placement, Liu et al. [19] set the transmission times of crossing data centers as an objective, Deng et al. [20] and Zhao et al. [21] targeted the data transmission volume, and Chen et al. [22] aimed for reducing the transmission costs. However, these methods did not involve the network bandwidth and its fluctuations, and thus, it is hard for them to map the data transmission time from their models to real-world network environments [23].

Moreover, most of traditional data placement strategies are based on deterministic environments. However, uncertainty is an essential feature of network environments, which may have a significant impact on data transmission [24]. Due to various loads between data centers, bandwidth fluctuations, network congestion, and other hardware characteristics, the data transmission time may be changeable even if the same data are transmitted between fixed data centers. Therefore, the uncertainty should be considered when building the data placement model for scientific workflows. In response to the uncertainty, the fuzzy theory has emerged as an effective tool [25]. Sun et al. [26] and Lei [27] fuzzified the processing time, completion time, and deadline, and then, the job scheduling method was studied under specific constraints. Based on the analytic hierarchy process (AHP) model, a data placement strategy was proposed in [28] to select the most suitable storage sites, which applied the fuzzy comprehensive evaluation to candidate data centers for different users. However, they did not involve the data placement problem of scientific workflows, and their fuzzy object and optimization goal was not the data transmission time.

To address the above problems, we proposed an effective data placement strategy for scientific workflows in hybrid cloud environments. The main contributions of this paper are summarized as follows:

(i) We define and model the data placement problem for scientific workflows in hybrid cloud environments. Specifically, we fuzzify the data transmission time into triangular fuzzy numbers and regard it as the optimization objective of the proposed model.

(ii) Based on the problem definitions and modeling, the DPSO-FGA is proposed as the second contribution for reducing the fuzzy data transmission time while considering the uncertainty of data transmission time, the different numbers and capacities of private data centers, and network bandwidth limitations, which can well adapt to real-world network environments.

(iii) We validate the effectiveness of the proposed DPSO-FGA method by using various scientific workflows in hybrid cloud environments, which can outperform the classic CFRA and CFGA methods in terms of fuzzy data transmission time.

The rest of this paper is organized as follows. Section 2 defines the data placement problem for scientific workflows in hybrid cloud environments. In Section 3, the proposed DPSO-FGA is discussed in detail. Section 4 shows the performance evaluation of the proposed method with simulation experiments. Finally, we conclude this paper and look for future work in Section 5.

2. Problem Definitions and Modeling

2.1. Problem Definitions

Definition 1. Hybrid cloud environment

A hybrid cloud environment consists of public and private data centers, where each private data center has a certain capacity, while each public data center has no capacity limitation. Thus, a hybrid cloud environment is defined as

$$\begin{cases} DC = \{DC_{\text{pub}}, DC_{\text{pri}}\}, \\ DC_{\text{pub}} = \{dc_1, dc_2, \ldots, dc_n\}, \\ DC_{\text{pri}} = \{dc_{n+1}, dc_{n+2}, \ldots, dc_{n+m}\}, \\ dc_i = \{V_i, \Delta_i\}, \end{cases} \tag{1}$$

where DC_{pub} is the set of public data centers, DC_{pri} is the set of private data centers, dc_i represents the i-th data center, and V_i indicates the maximum capacity of a data center. Specifically, the capacity of a public data center is unlimited, while a private data center may reserve some storage space with an upper limit V_i, and $\Delta_i \in \{0, 1\}$ represents the attribute of dc_i. If $dc_i \in DC_{\text{pub}}$, then $\Delta_i = 0$, and dc_i can be used to store public data. If $dc_i \in DC_{\text{pri}}$, then $\Delta_i = 1$, and dc_i can be used to store both public and private data. For any two data centers dc_i and dc_j, b_{ij} represents the network bandwidth between them, which is assumed to be known and fluctuate within a certain range.

Definition 2. Scientific workflow

The scientific workflow is a data-intensive application consisting of tasks and datasets, where a task may be related to multiple datasets and a dataset may also be related to multiple tasks. There is a data dependency relationship between the tasks, where the output datasets of a task may be the input datasets of other tasks. Meanwhile, there is also a

sequential relationship between the tasks, where a task may only be executed after all its predecessor tasks have been executed. After all the tasks are completed, the scientific workflow ends. In particular, the task without a predecessor task is the beginning task and the task without a successor task is the ending task. Moreover, datasets can be divided into initial and generated datasets, where the original input datasets of a scientific workflow are the initial datasets and the datasets generated during the running process are the generated datasets. Also, datasets can be divided into private and public datasets, where private datasets can only be stored in private data centers and the tasks using them as the input datasets must also be scheduled to the same data centers. By contrast, public datasets have no restriction on storage locations. Therefore, a scientific workflow is defined as a directed acyclic graph (DAG), denoted by G as

$$\begin{cases} G = \langle T, E, DS \rangle, \\ T = \{t_1, t_2, \ldots, t_c\}, \\ E = \{e_{12}, e_{13}, \ldots, e_{ij}\}, \\ DS = \{ds_1, ds_2, \ldots, ds_l\}, \\ t_i = \{I_i, O_i, DC(t_i)\}, \\ ds_i = \langle v_i, gt_i, lc_i \rangle, \end{cases} \tag{2}$$

where T is the set of tasks in G, E is the set of data dependencies between different tasks in G, and DS is the set of datasets in G. Specifically, t_c represents the c-th task and e_{ij} indicates the data dependency between tasks t_i and t_j, where $e_{ij} = 1$ indicates that t_i is the direct predecessor task of t_j. Moreover, ds_l is the l-th dataset, I_i is the input dataset of t_i, O_i is the output dataset of t_i, and $DC(t_i)$ is the data center for executing t_i. Furthermore, v_i is the size of the dataset ds_i, gt_i is the task number of generating ds_i, in which gt_i of the initial dataset is 0, and lc_i is the serial number of the data center storing ds_i.

It should be noted that the settings of privacy datasets in scientific workflows need to satisfy three logical rules. Specifically, for the task t_i in the hybrid cloud environment DC, when the set of input or output datasets of t_i (denoted by {Ii, Oi}) contains the privacy dataset ds_i, the following holds:

(i) Rule 1. $lc_i \equiv A$.

(ii) Rule 2. $DC(t_i) \equiv A$.

(iii) Rule 3. For each privacy dataset $ds_j \in \{I_i, O_i\}$, $lc_j \equiv DC(t_j) \equiv A$.

According to Definition 2, private datasets can only be stored in private data centers, while their storage locations cannot be changed. As shown in Rule 1, private datasets cannot be transmitted across data centers, and thus, the data center for executing a task using the dataset as an input or output must be fixed. As known from Rule 2, locations of private datasets for fixed tasks must be consistent with execution locations of the tasks, and thus, the privacy datasets cannot be stored in other data centers. Otherwise, the tasks cannot be executed.

Definition 3. Fuzzy data transmission time

When optimizing uncertainty problems, there are commonly three types of theories, including the probability theory, gray theory, and fuzzy theory. Specifically, the probability theory can be applied in sampling problems with massive samples, the gray theory is suitable for the problems with fewer samples, and the fuzzy theory can be used to solve the problems with unclear extensions of concepts [29]. As for addressing the data placement problem of scientific workflows, the fuzzy theory can be regarded as an effective tool because this problem has no clear boundary or the limitation involves uncertainty.

In the past research, the data transmission time was usually defined as the ratio of the dataset size to the bandwidth between data centers, without considering the other essential factors such as bandwidth fluctuations. However, the data transmission time is uncertain in real-world network environments. In response to this uncertainty, by utilizing the fuzzy theory, triangular fuzzy numbers are introduced to represent the data transmission time. For each independent data transmission process, the mapping $\langle dc_i, ds_k, dc_j \rangle$ indicates that the dataset ds_k is transmitted from the data center dc_i to dc_j. Therefore, the fuzzy data transmission time is defined as

$$\tilde{T}_{\text{transfer}}(dc_i, ds_k, dc_j) = (a^1_{ikj}, a^2_{ikj}, a^3_{ikj}), \tag{3}$$

where a^1_{ikj} and a^3_{ikj} are the lower and upper bound elements of the triangular fuzzy number, respectively. When $a^1_{ikj} = a^2_{ikj} = a^3_{ikj}$, the triangular fuzzy number degenerates into a real number. Moreover, the membership function indicates the degree which the element x belongs to the fuzzy interval. When $x = a^2_{ikj}$, the element x completely belongs to the interval. The membership function of the triangular fuzzy number is defined as

$$\mu = \begin{cases} \dfrac{x - a^1_{ikj}}{a^2_{ikj} - a^1_{ikj}}, & a^1_{ikj} \le x \le a^2_{ikj}, \\ \dfrac{x - a^3_{ikj}}{a^2_{ikj} - a^3_{ikj}}, & a^2_{ikj} \le x \le a^3_{ikj}, \\ 0, & \text{else.} \end{cases} \tag{4}$$

Definition 4. Calculation of fuzzy number

(1) The model involves addition and comparison operations between fuzzy numbers. For the triangular fuzzy numbers $\tilde{s} = (s_1, s_2, s_3)$ and $\tilde{t} = (t_1, t_2, t_3)$, the above operations are defined as follows:

(i) Addition operation (calculating the fuzzy data transmission time):

$$\tilde{s} + \tilde{t} = \tilde{t} + \tilde{s} = (s_1 + t_1, s_2 + t_2, s_3 + t_3). \tag{5}$$

(ii) Comparison operation (comparing the fuzzy completion time and choosing suitable values).

For $\widetilde{s} = (s_1, s_2, s_3)$, three comparison values are defined as

$$\begin{aligned} c_1(\widetilde{s}) &= \frac{(s_1 + 2s_2 + s_3)}{4}, \\ c_2(\widetilde{s}) &= s_2, \\ c_3(\widetilde{s}) &= s_3 - s_1. \end{aligned} \tag{6}$$

According to the literature [30], if $c_1(\widetilde{s}) > c_1(\widetilde{t})$, $\widetilde{s} > \widetilde{t}$. If $c_2(\widetilde{s}) > c_2(\widetilde{t})$, $\widetilde{s} > \widetilde{t}$; if $c_3(\widetilde{s}) > c_3(\widetilde{t})$, $\widetilde{s} > \widetilde{t}$; otherwise, $\widetilde{s} = \widetilde{t}$.

(2) The model involves the addition, subtraction, multiplication, division, fuzzification, and defuzzification operations between fuzzy and real numbers. For a triangular fuzzy number $\widetilde{s} = (s_1, s_2, s_3)$ and a real number t, the above operations are defined as follows:

(i) Addition and subtraction operations:

$$\begin{aligned} \widetilde{s} + t &= t + \widetilde{s} = (s_1 + t, s_2 + t, s_3 + t), \\ \widetilde{s} - t &= -(t - \widetilde{s}) = (s_1 - t, s_2 - t, s_3 - t). \end{aligned} \tag{7}$$

(ii) Multiplication and division operations:

$$\begin{aligned} \widetilde{s} \cdot t &= t \cdot \widetilde{s} = (s_1 t, s_2 t, s_3 t), \\ \frac{\widetilde{s}}{t} &= \left(\frac{s_1}{t}, \frac{s_2}{t}, \frac{s_3}{t}\right), \quad \text{where } t \neq 0. \end{aligned} \tag{8}$$

(iii) Fuzzification and defuzzification operations.

On the one hand, the fuzzification operation, according to the literature [26], is defined as

$$\begin{cases} s_1 \in [\delta_1 s, s], \\ s_2 = s, \\ s_3 \in [s, \delta_2 s], \end{cases} \tag{9}$$

where $\delta_1 < 1, \delta_2 > 1$.

On the other hand, the defuzzification operation is commonly used to quantitatively compare fuzzy numbers and analyze results. Li [31] defined the mean and standard deviation of fuzzy numbers under uniform distribution and proportional distribution, where the proportional distribution is suitable for the uncertainty problem of data transmission time. For the triangular fuzzy number $\widetilde{S}_{\text{total}}$, the mean and standard deviation are defined as

$$\begin{cases} \widetilde{S}_\mu = \dfrac{\int x\widetilde{S}^2_{\text{total}}(x)\mathrm{d}x}{\int \widetilde{S}^2_{\text{total}}(x)\mathrm{d}x} = \dfrac{s_1 + 2s_2 + s_3}{4}, \\ \widetilde{S}_\sigma = \left[\dfrac{\int x\widetilde{S}^2_{\text{total}}(x)\mathrm{d}x}{\int \widetilde{S}^2_{\text{total}}(x)\mathrm{d}x} - S^2_\mu\right]^{1/2} = \left[\dfrac{2(s_1 - s_2)^2 + (s_1 - s_3)^2 + 2(s_2 - s_3)^2}{80}\right]^{1/2}, \\ \min \widetilde{S}_{\text{total}} = (s_1, s_2, s_3) \overset{\text{translate}}{\Rightarrow} \min \widetilde{S}_\mu + \partial \widetilde{S}_\sigma, \partial \geq 0, \end{cases} \tag{10}$$

where $\widetilde{S}_\mu$ is the mean of the triangular fuzzy number $\widetilde{S} = (s_1, s_2, s_3)$, which reflects the most likely value of the fuzzy number under probability measurement. $\widetilde{S}_\sigma$ indicates the standard deviation of $\widetilde{S} = (s_1, s_2, s_3)$, which reflects the uncertainty degree of the fuzzy number. ∂ represents the weight of $\widetilde{S}_\sigma$.

Definition 5. Data placement strategy

The purpose of effective data placement is to reduce the data transmission time while meeting the order of task execution, the proportion of dataset privacy, and the capacity constraints of data centers. Only when the datasets required by a task are transmitted to the same data center, the task can be executed. Moreover, the time of scheduling a task to a data center is much shorter than the data transmission time [32], and thus, the model may need to focus on the data placement strategy. Before executing a task, the data center with the least fuzzy data transmission time will be chosen to schedule the task. Therefore, the data placement strategy is defined as

$$\begin{cases} S = (DS, DC, M, \widetilde{T}_{\text{total}}), \\ M = \underset{i=1,2,\ldots,|DC|}{\cup} \{dc_i, ds_k, dc_j\}, \\ \widetilde{T}_{\text{transfer}}(dc_i, ds_k, dc_j) = \left(a^1_{ikj}, a^2_{ikj}, a^3_{ikj}\right), \\ \widetilde{T}_{\text{total}} = \sum_{i=1}^{DC} \sum_{j \neq i}^{DC} \sum_{k=1}^{DS} \widetilde{T}_{\text{transfer}}(dc_i, ds_k, dc_j) \cdot e_{ijk}, \end{cases} \tag{11}$$

where M represents the mapping between the dataset DS and the set of data centers DC. $\{dc_i, ds_k, dc_j\}$ indicates that the dataset ds_k is transmitted from the data center dc_i to dc_j.

$\widetilde{T}_{\text{transfer}}(dc_i, ds_{\underline{k}}, dc_j)$ is the fuzzy data transmission time of $\{dc_i, ds_k, dc_j\}$. $\widetilde{T}_{\text{total}}$ is the total fuzzy data transmission time during the operation of scientific workflows, where $e_{ijk} = \{0,1\}$ indicates whether there is $\{dc_i, ds_k, dc_j\}$ during this time ($e_{ijk} = 1$ for yes and $e_{ijk} = 0$ for no).

2.2. Modeling. According to the above definitions, the data placement problem for scientific workflows is modeled based on the fuzzy theory with the objective of minimizing the fuzzy data transmission time while considering the capacity constraints of data centers. The problem model is defined as

$$\begin{cases} \min \widetilde{T}_{\text{total}} \\ \text{s.t. } \forall i, \sum_{j=1}^{DS} v_j \cdot u_{ij} \le V, \end{cases} \tag{12}$$

where $u_{ij} = \{0,1\}$ indicates whether the dataset ds_j is stored in the data center dc_i ($u_{ij} = 1$ for yes and $u_{ij} = 0$ for no).

3. Effective Data Placement for Scientific Workflows Based on DPSO-FGA

In light of the advantages of the particle swarm optimization (PSO), genetic algorithm (GA), and fuzzy theory, we propose an adaptive discrete particle swarm optimization algorithm based on the fuzzy theory and genetic algorithm operators (DPSO-FGA) to implement the effective data placement for scientific workflows, with the goal of minimizing the fuzzy data transmission time.

3.1. PSO Algorithm. The PSO algorithm was derived from the literature [33], which was inspired by the regularity in the cluster activities of flying birds. Based on the information exchanges between individuals, the movement of the entire population will gradually become orderly, and eventually, the optimal solution can be obtained, where the solution of the optimization problem is called "particle". When running the PSO algorithm, a fixed-size particle swarm is randomly initialized, where each particle constantly keeps iterating and updating through tracking the optimal solution found by itself and the population. The update of particles contains two aspects as follows.

(i) Speed update:

$$\begin{aligned} V_i(t+1) = wV_i(t) + c_1 r_1 [p_i(t) - X_i(t)] \\ + c_2 r_2 [g(t) - X_i(t)]. \end{aligned} \tag{13}$$

(ii) Location update:

$$X_i(t+1) = X_i(t) + V_i(t+1), \tag{14}$$

where the detailed definitions of the symbols can be found in [29].

3.2. Fitness Function. A fitness function needs to be defined for particles to track the optimal solution during the update process. As the optimization goal, the fuzzy data transmission time is used to define the fitness function as

$$F(S) = \widetilde{T}_{\text{total}}(X_i), \tag{15}$$

where $F(S)$ represents the fitness function of data placement strategy S and $\widetilde{T}_{\text{total}}(X_i)$ indicates the fuzzy data transmission time of the particle X_i.

If the total size of datasets placed in a data center does not exceed its maximum capacity, the particle can be a feasible solution. Otherwise, it is infeasible. When making the selection between feasible and infeasible solutions, the feasible one will be directly selected. When making the selection between feasible solutions, the particle with the smaller fitness function will be selected. When making the selection between infeasible solutions, the particle with the smaller fitness function will be also selected because it is more likely to become a feasible solution in subsequent operations.

3.3. Particle Encoding. The particle encoding needs to meet three principles, including completeness, nonredundancy, and soundness [34]. Specifically, the n-dimensional particles are discretely encoded [35], where n represents the number of datasets involved in the scientific workflow. Therefore, the structure of the particle i at the t-th iteration is defined as

$$X_i(t) = \left[x_i^1(t), x_i^2(t), \ldots, x_i^n(t)\right], \tag{16}$$

where $x_i^k(t)$ represents the placement location of the k-th dataset at the t-th iteration.

The following is an example of the particle encoding, where the particle number is 3, the current iteration number is 10, the number of datasets is 10, and the number of data centers is 4. Moreover, the datasets with underlines indicate that they are privacy datasets, where the data centers used for storing them cannot be changed during the subsequent update process:

$$X_3(10) = [1, 2, \underline{4}, 3, 2, \underline{1}, \underline{3}, 4, 2, 1]. \tag{17}$$

3.4. Particle Update. The traditional update of the PSO algorithm is shown in equations (13) and (14)fd16, which reveal some weaknesses in practical operation, such as low search ability, small solution space, and the premature convergence to the local optimum. To address these problems, crossover and mutation operations in the GA are introduced into the update. Since there are certain proportions of privacy datasets, the data centers used for storing them cannot be changed during the update process.

For the inertial part of the traditional update, the mutation operation is defined as

$$A_i(t+1) = \omega \oplus M_u(X_i(t)) = \begin{cases} M_u(X_i(t)), & r_0 < \omega, \\ X_i(t), & \text{else}, \end{cases} \tag{18}$$

where $r_0 \in (0,1)$ represents a random factor and the mutation operation $M_u()$ is to randomly change a quantile in the encoded particle within a range of values. It should be noted that the quantiles of privacy datasets cannot be mutated. Moreover, an infeasible particle should select a quantile, which makes the particle infeasible, to be mutated. Thus, the quantile to be selected should be the location of an overloaded data center. For instance, the following mutation operation is taken based on equation (17) as

$$\begin{aligned} &X_3(10) = [1, 2, \underline{4}, 3, 2, \underline{1}, \underline{3}, 4, 2, 1], \\ &\Downarrow \text{Mutation} \\ &A_3(11) = [1, 3, \underline{4}, 3, 2, \underline{1}, \underline{3}, 4, 2, 1], \end{aligned} \tag{19}$$

where the 2nd quantile is selected by the mutation operation and the corresponding data center number changes from 2 to 3.

For the individual and population cognitions in the traditional update, the cross operation is introduced as

$$\begin{aligned} B_i(t+1) &= c_1 \oplus C_p\left(A_i(t+1), p_i(t)\right) = \begin{cases} C_p\left(A_i(t+1), p_i(t)\right), & r_1 < c_1, \\ A_i(t+1), & \text{else}, \end{cases} \\ C_i(t+1) &= c_2 \oplus C_g\left(B_i(t+1), g(t)\right) = \begin{cases} C_g\left(B_i(t+1), g(t)\right), & r_2 < c_2, \\ B_i(t+1), & \text{else}, \end{cases} \end{aligned} \tag{20}$$

where $r_1, r_2 \in (0,1)$ represent random factors; $C_p(A_i(t+1), p_i(t))$ and $C_g(B_i(t+1), g(t))$, are to randomly select two quantiles of the encoded particles $A_i(t+1)$ and $B_i(t+1)$ and cross with the values at the same positions of $p_i(t)$ and $g(t)$. It should be noted that the storage locations of privacy datasets cannot be changed when crossing. For instance, the following crossover operation is taken based on (17) fd19 as

$$\begin{aligned} \frac{p_3(10)}{g(10)} = &\quad [1, 2, \underline{4}, 2, 4, \underline{1}, \underline{3}, 4, 2, 1], \\ &\downarrow\downarrow \text{Crossover} \\ \frac{A_3(11)}{B_3(11)} = &\quad [1, 2, \underline{4}, 2, 4, \underline{1}, \underline{3}, 4, 2, 1], \\ &\downarrow\downarrow \text{Turn Into} \\ \frac{B_3(11)}{C_3(11)} &\quad [1, 2, \underline{4}, 2, 4, \underline{1}, \underline{3}, 4, 2, 1], \end{aligned} \tag{21}$$

where the crossover operations happens at the 4th and 5th quintiles and the serial numbers of the data centers on the quantiles change from 2 to 4.

In summary, the process of the particle update is defined as

$$X_i(t+1) = c_2 \oplus C_g\Big(c_1 \oplus C_p\left(\omega \oplus M_u\left(X_i(t)\right), p_i(t)\right), g(t)\Big). \tag{22}$$

3.5. Mapping from Particles to Data Placement Results. Algorithm 1 shows the mapping from encoded particles to data placement results, where G represents the scientific workflow, DC indicates the hybrid cloud environment, X denotes the encoded particles, and S expresses the data placement strategy.

The execution steps of Algorithm 1 are listed as follows:

Step 1 (line 1). Initialize the current storage capacity of data centers $dc_{cur(i)}$ to 0 and the fuzzy data transmission time $\widetilde{T}_{\text{total}}$ to (0, 0, 0).

Step 2 (lines 2~7). Place each initial dataset into data centers according to their numbers and update $dc_{cur(i)}$. If $dc_{cur(i)}$ exceeds the maximum capacity of data centers, the solution corresponding to the particle is infeasible and the current operation will be stopped and returned.

Step 3 (lines 8~15). During the task traversal, the data center dc_j with the smallest fuzzy data transmission time will always be selected to place the task t_j. If the solution corresponding to the particle is infeasible (i.e., the sum of $dc_{cur(j)}$, $sum(I_j)$, and $sum(O_j)$ exceeds the maximum capacity of data centers), the current operation will be stopped and returned. Otherwise, the output dataset O_j of the task t_j will be placed into a corresponding data center and the storage capacity of data centers will be updated.

Step 4 (lines 16~20). Traverse all tasks, calculate the fuzzy data transmission time $\widetilde{T}_j$ for each dataset that needs to be transmitted across data centers, and sum them to obtain the total fuzzy data transmission time $\widetilde{T}_{\text{total}}$.

Step 5 (line 21). Output $\widetilde{T}_{\text{total}}$, X, and the corresponding data placement strategy.

3.6. Model Parameters. The inertial weight w in (13) has a direct influence on the convergence of the PSO algorithm [36], which can affect the search speed of particles in solution space. Thus, we propose a new method to define the weight w as follows, which can adaptively adjust the value of w based on the pros and cons of the corresponding solution of the current particle (i.e., the degree of difference between the current and historically optimal particles):

Procedure dataPlacement(G, DC, X)
Input: G, DC, X
Output: $S = (DS, DC, Map, \widetilde{T}_{\text{total}})$
1: **Initialization:** set the current storage capacity of data centers $dc_{cur(i)}$ to 0 and the fuzzy data transmission time $\widetilde{T}_{\text{total}}$ to (0, 0, 0)
2: **for each** ds_i **of** DS_{ini}//*Determine whether the particle would cause the data center overloaded*
3: $dc_{cur(X[i])} += v_i$//*Place the dataset ds_i in the data center $dc_{X[i]}$*
4: **if** $dc_{cur(X[i])} > V_{X[i]}$
5: return ***this particle is infeasible***
6: **end if**
7: **end for**
8: **for** $j = 1$ **to** $|T|$//*Determine whether the data center is overloaded during task execution*
9: Place the task t_j in the data center dc_j with the least fuzzy data transmission time
10: **if** $dc_{cur(j)} + sum(I_j) + sum(O_j) > v_j$
11: **return *this particle is infeasible***
12: **end if**
13: Place the output dataset O_j of t_j into the corresponding data center
14: Update the storage capacity of the data center
15: **end for**
16: **for** $j = 1$ **to** $|T|$//*Calculate the fuzzy transmission time of the corresponding data layout*
17: Find the data centers ds_i in the placement of task t_j's input dataset I_j
18: Calculate the fuzzy data transmission time $\widetilde{T}_j$ generated by the input dataset I_j to ds_j
19: $\widetilde{T}_{\text{total}} += \widetilde{T}_j$
20: **end for**
21: Output $\widetilde{T}_{\text{total}}$, $X$and the corresponding data placement strategy
End procedure

Algorithm 1: Mapping from encoded particles to data placement results.

$$w = w_{\max} - (w_{\max} - w_{\min}) \cdot \exp\left\{\frac{d(X_i(t), g(t))/|DS|}{(d(X_i(t), g(t))/|DS|) - 1.01}\right\}, \tag{23}$$

where $d(X_i(t), g(t))$ represents the degree of difference between the current particle $X_i(t)$ and the historically optimal particle $g(t)$ of the current population (i.e., the number of different values on the same quantiles). In the early stage of training, $d(X_i(t), g(t))$ is usually large with the large value of w. Therefore, it is necessary to expand the search range of particles in solution space, in order to find the optimal solution and avoid prematurely falling into local optimum. In the later stage of training, $d(X_i(t), g(t))$ becomes small with the small value of w. Thus, it is better to narrow the search range of particles and accelerate the search speed of particles for the optimal solution.

Moreover, the individual and population cognition factors (i.e., c_1 and c_2) are defined by using the gradient descent method [37].

4. Performance Evaluation

4.1. Parameter and Environment Settings. The scientific workflow model comes from five different scientific fields [38], including CyberShake, Epigenomics, Inspiral, Montage, and Sipht. Each scientific field has a scientific workflow with a different number of tasks, and each scientific workflow has a unique task structure, number of datasets, and computational requirements [39]. Specifically, a medium-sized (about 50 tasks) workflow in each field is selected for experiments, and the parameter and environment settings are shown in Table 1.

Moreover, some extrasettings are shown as follows:

(i) Maximum capacity: the datum capacity is set to $\sum_{i=1}^{|DS|} v_i / (|DC| - 1)$, where the maximum capacity of the three private data centers is set to 2.6 times the datum capacity.

(ii) Bandwidth (M/s) between data centers: the bandwidth between dc_1 and $\{dc_2, dc_3, dc_4\}$ is set to {10, 20, 30}, the bandwidth between dc_2 and $\{dc_3, dc_4\}$ is set to {150, 150}, and the bandwidth between dc_3 and dc_4 is set to 100.

(iii) Proportion of privacy datasets: due to the difference of datasets among various workflows, the proportions of privacy datasets of in the scientific workflows, including CyberShake, Epigenomics, Inspiral, Montage, and Sipht, are set to {0.25, 0.2, 0.2, 0.2, 0.02}.

(iv) Fuzzy parameter: based on the fuzzy theory, the data transmission time T is fuzzified into the triangular fuzzy number $\widetilde{T}$, where the fuzzy parameters are set to $\sigma_1 = 0.85, \sigma_2 = 1.2$.

4.2. Comparison Algorithms. The proposed DPSO-FGA is compared with the constraint fuzzy randomized algorithm (CFRA) and constraint fuzzy greedy algorithm (CFGA), which can improve the performance of the randomized algorithm (RA) and greedy algorithm (GA) in data placement. The CFRA and CFGA rely on the fuzzy theory while considering some essential conditions, including the application scenarios of scientific workflows, privacy settings, and capacity constraints. The conditions refer to meet the maximum capacity

Table 1: Environment and parameter settings.

Parameter	Value
Population size	100
Maximum number of iterations	1000
$w_{\max}$, $w_{\min}$	0.9, 0.4
$c_1^{\text{start}}, c_1^{\text{end}}$	0.9, 0.2
$c_2^{\text{start}}, c_2^{\text{end}}$	0.4, 0.9
Hybrid cloud environment	$DC = \{DC_{\text{pub}}, DC_{\text{pri}} \mid DC_{\text{pub}} = \{dc_1\}, DC_{\text{pri}} = \{dc_2, dc_3, dc_4\}\}$
Software environment	Windows 10 64-bit, Python 3.7
Hardware environment	Intel® Core™ i7-6700HQ CPU @2.60GHZ, RAM 8.00 GB

requirements of data centers and the proportion of private datasets during the data placement process.

(i) The steps of CFRA

Step 1. Set privacy datasets and the maximum capacity of data centers, initialize parameters, and keep the same values as the corresponding parameters in the DPSO-FGA.

Step 2. Generate a random population that meets the conditions according to the discrete encoding method in the DPSO-FGA. The population contains a certain number of individuals, and each individual represents a candidate solution for data placement.

Step 3. Define the fitness function as the fuzzy data transmission time of the corresponding solution of the individual encoding.

Step 4. Calculate the fitness value of each individual and compare it with the current best individual of the population. If the current individual is the better one, update the best individual of the population by using the current one.

Step 5. End the traversal and output the best individual with its fitness value.

(ii) The steps of CFGA

Step 1. Set privacy datasets and the maximum capacity of data centers, initialize parameters, and keep the same values as the corresponding parameters in the DPSO-FGA.

Step 2. Design a data placement strategy. According to the task execution sequence of the scientific workflow, traverse the datasets that have not been deployed for tasks. If the current task has been placed, the dataset will also be placed to the same data center by using the GA. If the current task has not been placed but there is already a placed dataset, the dataset will be placed to the same data center as the placed dataset by using the GA. If the current task has not been placed and there is no already placed dataset, the dataset will be placed into the data center with the smallest fuzzy data transmission time by using the GA.

Step 3. Calculate the fuzzy data transmission time of the current data placement strategy and output the strategy.

4.3. Experimental Results and Analysis. To avoid the randomness of results, 10 independent experiments are carried out on five scientific workflows under different environment settings. Table 2 records the average fuzzy data transmission time of different algorithms under various scientific workflows.

In the subsequent experimental results, the fuzzy data transmission time is defuzzified, in order to make the comparison between the algorithms more intuitive, where ∂ is set to 1.

Figure 1 shows the defuzzification results for the fuzzy data transmission time of different algorithms under various scientific workflows, where the names of the scientific workflows are indicated by their first letter.

From the perspective of algorithms, the DPSO-FGA outperforms the CFRA and CFGA. This is because that the CFGA may easily fall into the local optimum by using the GA during execution, and thus, it ignores the global performance. Moreover, the overall performance of the CFRA is better than CFGA since the search space of the CFRA is larger than the CFGA and will not fall into the local optimum, and thus, the CFRA can obtain a good solution when the algorithm runs for a long time. However, the CFRA does not consider the fitness of the current particle when a solution is generated, and thus, the performance of the CFRA is worse than the DPSO-FGA. From the perspective of workflows, the data transmission time of the same algorithm in various scientific workflows is significantly different. Although all these scientific workflows contain about 50 tasks, the number of datasets used varies greatly. For example, the CyberShake uses datasets only about 70 times, while the Sipht uses datasets up to 4000 times, which results in the different data transmission time between them.

As the number of private data centers in a hybrid cloud environment sometimes changes, the performance of DPSO-FGA needs to be evaluated with different numbers of private data centers. Thus, we change the number of private data centers without modifying other default settings. Specifically, these three algorithms are tested when the number of private data centers is set to {3, 5, 6, 8, 10}, where the bandwidth between newly added private data centers and public data centers is set to 20 M/s and the bandwidth between other private data centers is set to 120 M/s. The experimental results are shown in Figure 2.

From the perspective of algorithms, the performance of the DPSO-FGA outperforms the CFRA and CFGA, and the reasons have been analyzed in Figure 1. From the perspective of private data centers, as the number of private data centers

Table 2: Average fuzzy data transmission time of different algorithms under various scientific workflows.

Workflow	Algorithm	Average fuzzy data transmission time (s)
CyberShake	DPSO-FGA	(375683.37, 415977.97, 448166.12)
	CFRA	(613685.49, 658202.55, 728115.64)
	CFGA	(760212.74, 819721.35, 890377.31)
Epigenomics	DPSO-FGA	(297570.10, 300909.35, 326971.76)
	CFRA	(561274.26, 606573.37, 654489.23)
	CFGA	(1192066.03, 1242136.17, 1290256.56)
Inspiral	DPSO-FGA	(2356942.36, 2521230.04, 2734877.94)
	CFRA	(5859201.90, 6335216.52, 7023988.75)
	CFGA	(8025939.27, 8547697.91, 9737463.34)
Montage	DPSO-FGA	(794955.76, 840543.65, 863448.55)
	CFRA	(1242644.26, 1331192.71, 1445410.25)
	CFGA	(1920877.78, 2058800.54, 2311869.58)
Sipht	DPSO-FGA	(10733197.95, 12051002.69, 1289802.65)
	CFRA	(11543263.40, 12494008.57, 13725762.53)
	CFGA	(13105310.17, 14687168.03, 1591557.53)

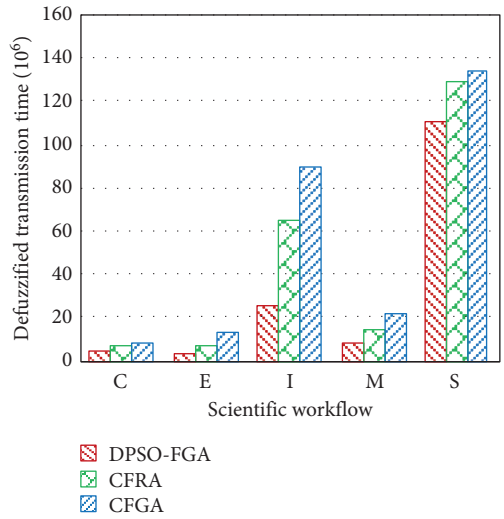


Figure 1: Average fuzzy data transmission time of different algorithms under various scientific workflows.

increases, the data transmission time of these three algorithms also inclines. This is because, with the increasing number of private data centers, the privacy datasets, which are randomly set according to the privacy proportion, are dispersed and fixed in more private data centers. Therefore, the fixed tasks, which require these private datasets, need to be executed in more scattered locations, and it will lead to increasing data transmission time.

Since the maximum capacity of private data centers is regarded as a constraint, the sensitivity of the DPSO-FGA to this constraint needs to be evaluated. Specifically, the CyberShake is selected as the scientific workflow for experiments, the multiple of datum capacity is set to {2, 2.6, 3, 5, 8}, and the rest of the settings remain default. The experimental results are shown in Figure 3.

When the maximum capacity of private data centers increases and the bandwidth between data centers remains the same, each data center is able to store more datasets and the datasets required for executing tasks become more concentrated. Therefore, the data transmission time of the DPSO-FGA is reduced. Specifically, the fastest decline in data transmission time happens when the maximum capacity is 2~3 times than that of the datum capacity, and the slowest decline happens when the maximum capacity is 5~8 times than that of the datum capacity. This is because when the maximum capacity of data centers is small, the available space is small and the placement locations of datasets are restricted. Thus, the maximum capacity has a significant impact on data transmission time. When the maximum capacity of data centers becomes larger, each data center can store more datasets, and it will be easy to meet the operational requirements of scientific workflows. Therefore, the maximum capacity has little effect on the data transmission time.

Finally, the performance of the DPSO-FGA is evaluated under different bandwidths between data centers. Specifically, the CyberShake is selected as the scientific workflow for experiments, the multiple of the bandwidth between data centers relative to the default one is set to {0.5, 0.8, 1.5, 3, 5}, and the rest of the settings remains default. The experimental results are shown in Figure 4. The data transmission time

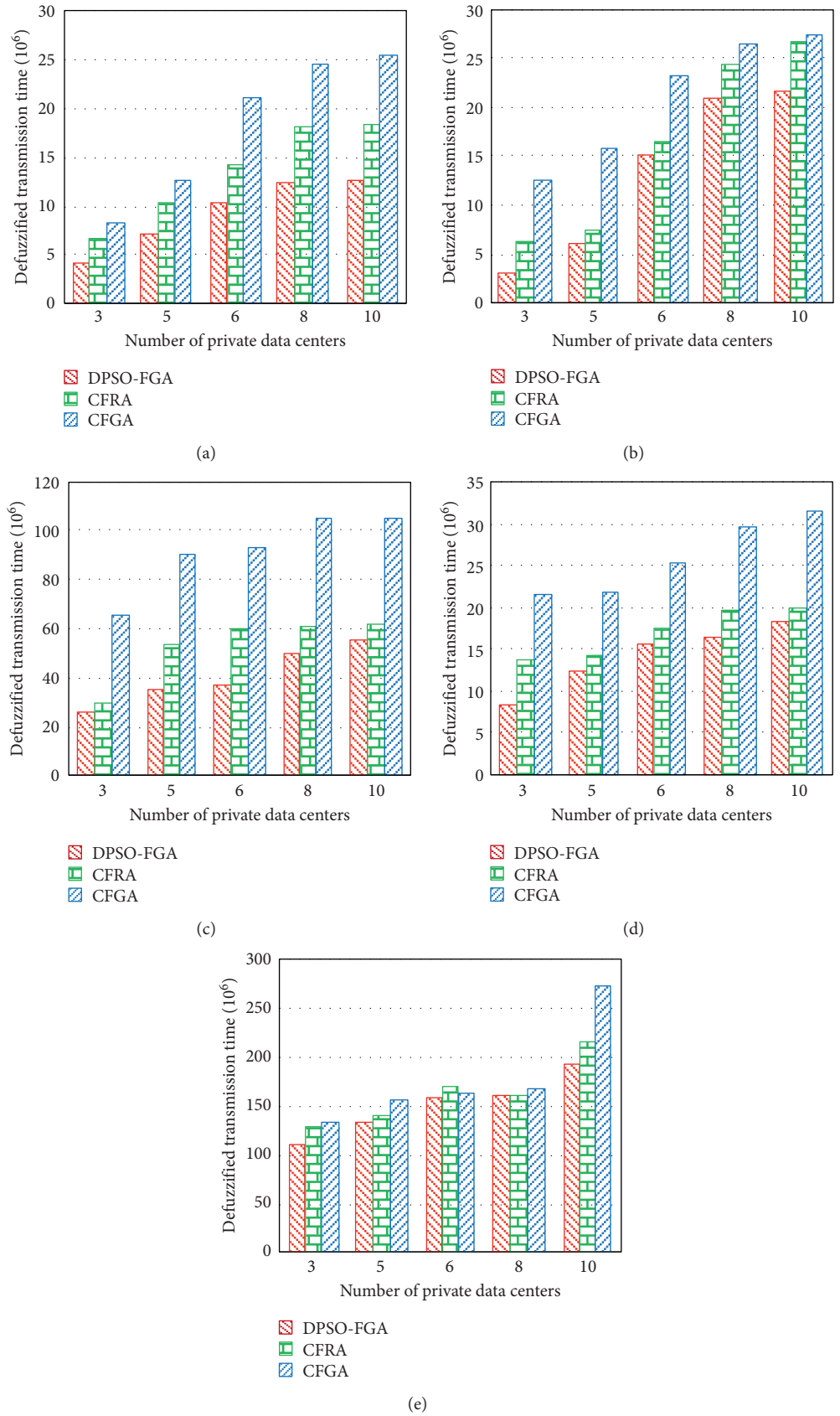


Figure 2: Average data transmission time of different algorithms with various numbers of private data centers: (a) Cybershake; (b) Epigenomics; (c) Inspiral; (d) Montage; (e) Sipht.

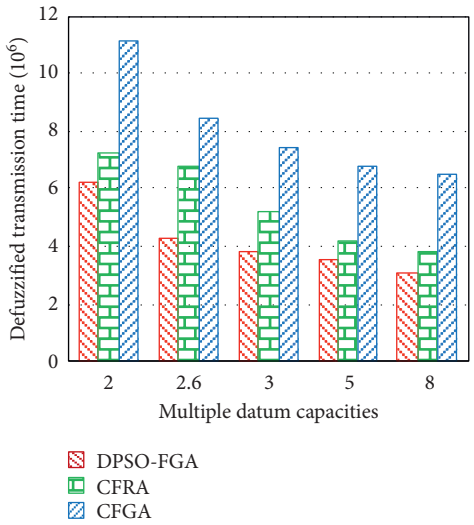


Figure 3: Average fuzzy data transmission time of different algorithms with various capacities of private data centers.

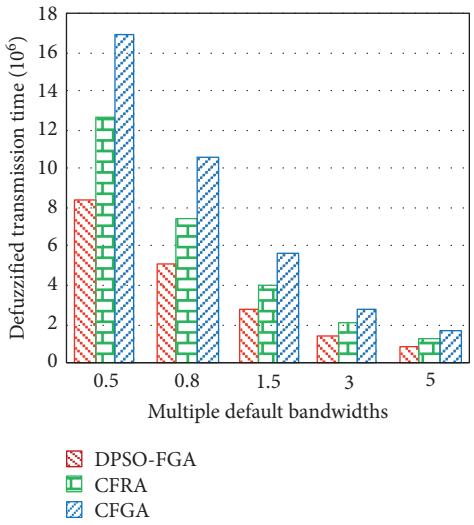


Figure 4: Average fuzzy data fuzzy transmission time of different algorithms with various bandwidths between data centers.

decreases greatly as the bandwidth increases, which indicates that the bandwidth changes between data centers will not significantly affect the data placement strategy.

5. Conclusions

In this paper, we propose a DPSO-FGA-based data placement method for scientific workflows in hybrid cloud environments. Based on the fuzzy theory, the DPSO-FGA fuzzifies the data transmission time for adapting to real-world network environments while considering the characteristics of hybrid cloud environments, bandwidth fluctuations, capacity limitations of private data centers, and dependencies between different scientific workflow tasks. Simulation results demonstrate the effectiveness of the proposed DPSO-FGA method. In the future, we will study the impact of other essential factors on the proposed method, such as different proportions of private datasets in scientific workflows and capacities of various private data centers. Moreover, under the conditions that are not critical for data transmission time, such as business network environments, data transmission costs between different clouds should also be regarded as a prioritized optimization goal. Therefore, a comprehensive model for minimizing the fuzzy data transmission time and costs will be researched.

Data Availability

The data used to support the findings of this study are produced by a public workflow generator available at https://confluence.pegasus.isi.edu/display/pegasus/WorkflowGenerator.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Authors' Contributions

Zheyi Chen and Xu Zhao are equally contributed to this work. Zheyi Chen and Xu Zhao developed the model, carried out the parameter estimations, and planned as well as performed the experiments. Zheyi Chen wrote the main part of the manuscript, while Xu Zhao and Bing Lin provided the support for writing materials. Bing Lin also took part in the design and evaluation of the model. Zheyi Chen and Bing Lin reviewed the manuscript. All the authors read and approved the final manuscript.

Acknowledgments

This work was supported by the National Key R&D Program of China (Grant no. 2018YFB1004800), Natural Science Foundation of China (Grant nos. 61672159, 41801324, and 61972165), Natural Science Foundation of Fujian Province (Grant nos. 2019J01286, 2019J01244, and 2018J01619), Young and Middle-Aged Teacher Education Foundation of Fujian Province (Grant no. JT180098), Open Foundation of Engineering Research Center of Big Data Application in Private Health Medicine, Fujian Province University (Grant no. KF2020001), Talent Program of Fujian Province for Distinguished Young Scholars in Higher Education, and China Scholarship Council (no. 201706210072). The authors sincerely thank Dr. Jia Hu and Dr. Geyong Min for providing useful advice that greatly improved this paper.

References

[1] B. Varghese and R. Buyya, "Next generation cloud computing: new trends and research directions," *Future Generation Computer Systems*, vol. 79, pp. 849–861, 2018.

[2] Z. Chen, J. Hu, G. Min, A. Y. Zomaya, and T. El-Ghazawi, "Towards accurate prediction for high-dimensional and highly-variable cloud workloads with deep learning," *IEEE Transactions on Parallel and Distributed Systems*, vol. 31, no. 4, pp. 923–934, 2020.

[3] X. Chen, F. Zhu, Z. Chen et al., "Resource allocation for cloud-based software services using prediction-enabled feedback control with reinforcement learning," *IEEE Transactions on Cloud Computing (TCC)*, 2020.

[4] K. Peng, H. Huang, S. Wan et al., "End-edge-cloud collaborative computation offloading for multiple mobile users in heterogeneous edge-server environment," *Wireless Networks*, 2020.

[5] X. Chen, H. Wang, Y. Ma, X. Zheng, and L. Guo, "Self-adaptive resource allocation for cloud-based software services based on iterative QoS prediction model," *Future Generation Computer Systems*, vol. 105, pp. 287–296, 2020.

[6] D. S. Linthicum, "Emerging hybrid cloud patterns," *IEEE Cloud Computing*, vol. 3, no. 1, pp. 88–91, 2016.

[7] B. Zhang, L. Yu, Y. Feng, L. Liu, and S. Zhao, "Application of workflow technology for big data analysis service," *Applied Sciences*, vol. 8, no. 4, Article ID 591, 2018.

[8] M. Atkinson, S. Gesing, J. Montagnat, and I. Taylor, "Scientific workflows: past, present and future," *Future Generation Computer Systems*, vol. 75, pp. 216–227, 2017.

[9] X. Li, L. Zhang, Y. Wu, X. Liu, E. Zhu et al., "A novel workflow-level data placement strategy for data-sharing scientific cloud workflows," *IEEE Transactions on Services Computing*, vol. 12, no. 3, pp. 370–383, 2019.

[10] B. Lin, F. Zhu, J. Zhang, J. Chen, X. Chen et al., "A time-driven data placement strategy for a scientific workflow combining edge computing and cloud computing," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 7, pp. 4254–4265, 2019.

[11] W. Guo, B. Lin, G. Chen, Y. Chen, and F. Liang, "Cost-driven scheduling for deadline-based workflow across multiple clouds," *IEEE Transactions on Network and Service Management*, vol. 15, no. 4, pp. 1571–1585, 2018.

[12] B. Lin, W. Guo, N. Xiong, G. Chen, A. V. Vasilakos et al., "A pretreatment workflow scheduling approach for big data applications in multicloud environments," *IEEE Transactions on Network and Service Management*, vol. 13, no. 3, pp. 581–594, 2016.

[13] D. Yuan, Y. Yang, X. Liu, and J. Chen, "A data placement strategy in scientific cloud workflows," *Future Generation Computer Systems*, vol. 26, no. 8, pp. 1200–1214, 2010.

[14] K. H. K. Reddy, V. Pandey, and D. S. Roy, "A novel entropy-based dynamic data placement strategy for data intensive applications in Hadoop clusters," *International Journal of Big Data Intelligence*, vol. 6, no. 1, pp. 20–37, 2019.

[15] X. Li, Y. Wu, X. Liu et al., "Datacenter-oriented data placement strategy of workflows in hybrid cloud," *Journal of Software*, vol. 27, no. 7, pp. 1861–1875, 2016, in Chinese.

[16] Z. Liu, T. Xiang, B. Lin et al., "A data placement strategy for scientific workflow in hybrid cloud," in *Proceedings of IEEE International Conference on Cloud Computing (CLOUD)*, pp. 556–563, San Francisco, CA, USA, July 2018.

[17] P. Zheng, L.-Z. Cui, H.-Y. Wang, and M. Xu, "A data placement strategy for data-intensive applications in cloud," *Chinese Journal of Computers*, vol. 33, no. 8, pp. 1472–1480, 2010, in Chinese.

[18] L. Cui, J. Zhang, L. Yue, Y. Shi, H. Li, and D. Yuan, "A genetic algorithm based data replica placement strategy for scientific applications in clouds," *IEEE Transactions on Services Computing*, vol. 11, no. 4, pp. 727–739, 2018.

[19] S.-W. Liu, L.-M. Kong, K.-J. Ren, J.-Q. Song, K.-F. Deng, and H.-Z. Leng, "A two-step data placement and task scheduling strategy for optimizing scientific workflow performance on cloud computing platform," *Chinese Journal of Computers*, vol. 34, no. 11, pp. 2121–2130, 2011, in Chinese.

[20] K. Deng, K. Ren, M. Zhu et al., "A data and task co-scheduling algorithm for scientific cloud workflows," *IEEE Transactions on Cloud Computing (TCC)*, Early Access, vol. 8, no. 2, pp. 349–362, 2020.

[21] Q. Zhao, C. Xiong, X. Zhao et al., "A data placement strategy for data-intensive scientific workflows in cloud," in *Proceedings of IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID)*, pp. 928–934, Shenzhen, China, May 2015.

[22] Z. Chen, J. Hu, G. Min, and X. Chen, "Effective data placement for scientific workflows in mobile edge computing using genetic particle swarm optimization," *Concurrency and Computation: Practice and Experience (CCPE)*, Article ID e5413, 2019.

[23] B. Lin, T. Xiang, G. Chen et al., "Time-driven data placement strategy for scientific workflows in hybrid cloud," *Computer Integrated Manufacturing Systems*, vol. 25, no. 4, pp. 909–919, 2019, in Chinese.

[24] A. Vafamehr, M. E. Khodayar, S. D. Manshadi, I. Ahmad, and J. Lin, "A framework for expansion planning of data centers in electricity and data networks under uncertainty," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 305–316, Jan. 2019.

[25] Z. Kovacic and S. Bogdan, *Fuzzy Controller Design: Theory and Applications*, CRC Press, Boca Raton, FL, USA, 2018.

[26] L. Sun, L. Lin, M. Gen, and H. Li, "A hybrid cooperative coevolution algorithm for fuzzy flexible job shop scheduling," *IEEE Transactions on Fuzzy Systems*, vol. 27, no. 5, pp. 1008–1022, 2019.

[27] D. Lei, "Fuzzy job shop scheduling problem with availability constraints," *Computers & Industrial Engineering*, vol. 58, no. 4, pp. 610–617, 2010.

[28] X. Wu and H. Guan, "Data set replica placement strategy based on fuzzy evaluation in the cloud," *Journal of Intelligent & Fuzzy Systems*, vol. 31, no. 6, pp. 2859–2868, 2016.

[29] X. Zhou, G. Zhang, J. Sun, J. Zhou, T. Wei, and S. Hu, "Minimizing cost and makespan for workflow scheduling in cloud using fuzzy dominance sort based HEFT," *Future Generation Computer Systems*, vol. 93, pp. 278–289, 2019.

[30] M. Sakawa and R. Kubota, "Fuzzy programming for multi-objective job shop scheduling with fuzzy processing time and fuzzy due date through genetic algorithms," *European Journal of Operational Research*, vol. 120, no. 2, pp. 393–407, 2000.

[31] R. Li, *Theory and Application of Fuzzy Multi-Criteria Decision*, Science Press, Beijing, China, 2002, in Chinese.

[32] K. Deng, K. Ren, J. Song et al., "A clustering based coscheduling strategy for efficient scientific workflow execution in cloud computing," *Concurrency and Computation: Practice and Experience (CCPE)*, vol. 25, no. 18, pp. 2523–2539, 2014.

[33] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of IEEE International Conference on Neural Networks (ICNN)*, pp. 1942–1948, Perth, Australia, November-December 2002.

[34] J. Su, W. Guo, C. Yu et al., "Fault-tolerance clustering algorithm with load-balance aware in wireless sensor network," *Chinese Journal of Computers*, vol. 37, no. 2, pp. 445–456, 2014, in Chinese.

[35] M. A. Rodriguez and R. Buyya, "Deadline based resource provisioningand scheduling algorithm for scientific workflows on clouds," *IEEE Transactions on Cloud Computing*, vol. 2, no. 2, pp. 222–235, 2014.

[36] M. Masdari, F. Salehi, M. Jalali, and M. Bidaki, "A survey of PSO-based scheduling algorithms in cloud computing,"

Journal of Network and Systems Management, vol. 25, no. 1, pp. 122–158, 2017.

[37] Y. Shi and R. Eberhart, "A modified particle swarm optimizer," in *Proceedings of IEEE International Conference on Evolutionary Computation (ICEC)*, pp. 69–73, Anchorage, AK, USA, May 1998.

[38] S. Bharathi, A. Chervenak, E. Deelman et al., "Characterization of scientific workflows," in *Proceedings of the IEEE Workshop on Workflows in Support of Large-Scale Science*, pp. 1–10, Austin, TX, USA, November 2008.

[39] "Workflow generator," 2014, https://confluence.pegasus.isi.edu/display/pegasus/WorkflowGenerator.

Hindawi
Discrete Dynamics in Nature and Society
Volume 2020, Article ID 1057902, 11 pages
https://doi.org/10.1155/2020/1057902

Research Article

Granularity Decision of Microservice Splitting in View of Maintainability and Its Innovation Effect in Government Data Sharing

Yan Li,[1] **Chun-Zi Wang**,[1] **Ying-chao Li**,[1,2] **and Jia Su**[3]

[1]*School of Management, Xi'an Polytechnic University, Xi'an 710048, China*
[2]*Emaplink Software Co., Ltd., Xi'an 710000, China*
[3]*School of Management, Xi'an University of Architecture & Technology, Xi'an 710055, China*

Correspondence should be addressed to Yan Li; sayidli@xpu.edu.cn

Received 18 May 2020; Revised 23 June 2020; Accepted 16 July 2020; Published 4 August 2020

Guest Editor: Chi-Hua Chen

The reasonable sharing and effective integration of data is the premise of improving the comprehensive governance ability of society through data empowerment. It can be determined that the fine-grained splitting of microservice framework can promote the reasonable sharing and effective integration of data. However, how to determine the granularity of microservice splitting is a multiparameter and multiobjective decision-making problem, which is also a key basic problem to be solved urgently both in academic research and application. From the perspective of application, this paper puts forward the criteria and basic framework that can guide the microservice splitting, and for the first time, based on the perspective of maintainability, it gives the decision criteria and methods of microservice splitting granularity. After the theoretical research, this paper also takes the provincial microservice governance of food and drug regulation as an example: the example shows that the framework and methods proposed can effectively improve data sharing and system expandability. Through microservice governance, collaborative social governance featuring "one network to achieve management objectives, one network to realize comprehensive business processing, one network for comprehensive view of all information" can be achieved. It is an exemplary mode worth popularization.

1. Introduction

As digitization evolves, humans are entering a finely resolved society, also known as the granular society [1] in which everyone must go through the transformation from "a rational person" to "a granular person." Technical innovation provides high-density and more detailed insight into all phenomena, and the high resolving process of digitization will also drive social system disintegration. As the result of this "resolving-disintegration" dual processes, population average value and outdated knowledge of the coarse-particle time will be phased out and the consequent "big-data mindset" will revolutionize the traditional philosophy of social governance. Thus, amid the granular time, how to innovate governance mode and modernize relative ability of the government is a significant issue urging for a solution.

At present, the world is witnessing an upsurging trend of rapid transformation from data into information and knowledge and the improvement of data-driven government governance ability [2, 3]. As China's reform enters the deep water zone, potential social contradictions and conflicts are looming. Thus, data-based speaking, decision, management, and innovation have become an effective way to enhance our comprehensive and collaborative social governance ability. On August 19, 2015, the State Council executive meeting adopted the Action Outline for Promoting Big Data Development, which defined how to transform government functions and optimize service modes through data sharing and integration from the national top-layer policy design level. Implementation Plan for the Integration and Sharing of Government Information System (General Office of the State Council [2017] No. 39) requires the combination of investigation and elimination to speed up zombie

information system removal and the integration and sharing of the internal information system of each department. On this basis, Notice on the Implementation of Government Information System Integration and Application Pilots (Department of High Technology Industry of National Development and Reform Commission [2017] No. 1714) issued by the National Development and Reform Commission also mentioned a couple of times that the foundation of information system integration and sharing is the integration of government data and the establishment of a unified sharing platform. This pointed out where the transformation and upgrading of China's e-government service ecosystem should go. Data integrated by e-government accounts for over 90% of the total amount of social data [4]. As the granular mode increasingly intensifies, data sharing and integration become the basic requirements for the synergy and accuracy of comprehensive social governance and represent a consensus reached by the academic and the industrial communities. At present, government data sharing and integration are mainly implemented through enterprise service bus (ESB) [3, 5, 6] and data center sharing [4, 7]. As a key component of service-oriented architecture (SOA), ESB gathers middleware, XML (Extensible Markup Language), Web services, and other technologies to support intra- or interenterprise heterogeneous systems in service, messaging, or event interactions. With standard open protocols and a proven application integration model, it acts as a connection hub linking various application systems for data sharing. However, responsiveness and maintenance cost are where its development constraints lie. The exchange service configuration is complex and users have to shoulder much workload. Service updates can only be achieved by reconstructing the original system, cleaning data, and transferring them to the terminal. These defects show that ESB can hardly meet exponentially growing interaction demands between information systems. The concept of data center sharing is to build a large centralized data resource pool level by level and realize data sharing between different organs through the data resource service catalog. The advantage is that it can form a unified data source and ensure data consistency, but this traditional data collecting method will inevitably involve data collection, comparison, cleaning, and heterogeneous data conversion of different organs and organizations, which makes real-time data hard to guarantee. Considering the authority and responsibility of departments, information sharing faces resistance and obstacles in large areas. Government data integration is greatly limited in many ways including information degradation in the two-way funnel filtering process of the upper and the lower levels in departments, information island caused by the interdepartment chimney system, and information obstruction at the terminal end led by the information asymmetry between the government and the public [7, 8].

Data sharing is the precondition of realizing the mutual function of big data and government governance. Therefore, government data sharing is a fundamental and key issue for both academic research and practical application. This paper builds an underlying platform of data sharing and service connection based on the distributed microservice architecture; the overall chapter structure is as follows. The first section is the introduction; the second section summarizes the history of software technology architecture; and on this basis, the third section clarifies the definition and advantages of microservice. In the fourth section, the paper introduces the framework and key technologies of the distributed microservice platform and for the first time discusses quantitative measurement indexes of microservice architecture. In the fifth section, the availability of the architecture is verified by the example of the collaborative social governance system of Shaanxi Medical Products Administration. The sixth section is the summary of the whole paper.

The main innovations and contributions of this paper are as follows: (1) systematically summarizes the differences and advantages between microservice architecture and other three mainstream architectures and clarifies the main application scenarios and functions of microservice; (2) for the first time, based on the perspective of maintainability, gives the decision-making standards and methods of microservice resolution granularity; and (3) from the perspective of application effect, takes the provincial food and drug supervision microservice governance. An example is given to demonstrate the significance of the proposed criteria and methods.

2. The Characteristics and Splitting Granularity of Microservices

As a systematic sketch, software architecture (SA) is the foundation of software practices and the series of principles and goals set by software systems. Though the concept is simple, it is hard to draw an accurate definition, even a descriptive one, so most engineers comprehend it from the intuitive perspective. Based on years of experience in software industry, this paper divides the evolution process of SA into monolithic, vertical, service-oriented, and microservice architectures.

Monolithic architecture (Figure 1) has no unified definition. Essentially, it organizes all basic applications into one module or unit [8] or, in other words, gathers all functions into a project during the development process and packs them into a WAR for further server deployment. Both time- and cost-saving, it is simple in structure and suitable for small-scale projects. Little interface interaction makes it relatively easy to optimize and upgrade execution performance of the system. However, disadvantages also exist: (1) all functions in one project impede the mutual expansion and maintenance work of multiple collaborators; (2) generally, system upgrade must rely on the clustering on both application and database levels, which is costly; and (3) mass data reading and writing technologies are still facing bottlenecks.

Derived from the concept of the vertical structure in nature (special vertical differentiation or stratification of communities), vertical architecture targets at the limitation of monolithic application architecture. As project complexity and the amount of users grew, the acceleration driven by cluster optimization in single applications tapered; thus,

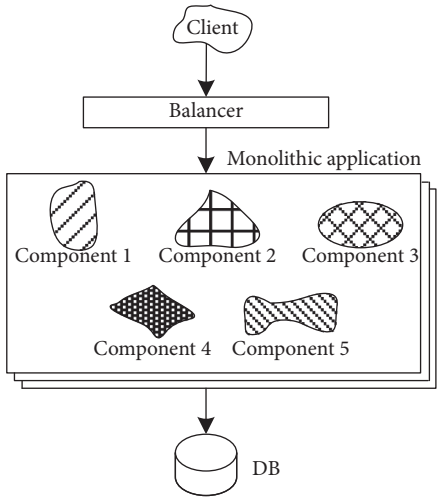


Figure 1: Framework of monolithic architecture.

people began to raise efficiency by vertically dividing large projects and splitting them into multiple monolithic architecture projects (Figure 2). Vertical architecture can, to some extent, alleviate the defects of traditional monolithic architecture. For example, vertical separation can resist the original monolithic project from excessively expanding, and different technologies can be used on different vertical projects, making collaboration and maintenance easy to operate by multiple people. But at root, it is still a continuation of monolithic architecture, so the essential problems have yet to be solved.

With the boom of vertical architecture applications, the interaction problem between different applications became salient. SOA gained popularity around 2000 (Figure 3). Accurately, it is an architecture style [9, 10]. In this service structure featuring coarseness and loose coupling, communication between services is achieved through simple and precisely defined interfaces, involving no underlying programmatic interfaces and communication models. Fundamental SOA consists of service provider, consumer, and registration. It abstracts repeatedly shared functions into modules and provides services for each system in the way of service. Through ESB, multiple services in the project communicate by means of Web service and RPC. The advantages of SOA are apparent: (1) the extraction of shared modules raises reusability and maintainability, thus promoting efficiency; (2) since ESB lowers the coupling degree between interfaces, different projects and services can adopt different technologies and clustering and the optimization plan can be tailored due to the characteristics of a certain service. Meanwhile, the vague boundary between system and service hinders SOA development and maintenance. Though with ESB, the unfixed and various service interface agreements still impede system maintenance. As a result, services extracted are too coarse, and the coupling degree of system and service is high.

As a new architecture concept put forward after 2010, microservice architecture has no unified definition at present [11–13]. The most representative one is proposed by Martin Fowler [12, 14] who thinks MSA is a specific designing method transforming software applications into independent and deployable service modules. It is structured in such a style that single applications are developed due to service abilities into microservices independently running in their own processes (Figure 4). Service communication is realized on the basis of lightweight protocols (such as RESTful), and each service can be independently deployed by fully automated mechanisms. This method, which splits out the service layer and extracts it into multiple small-scale services, has obvious advantages. Compared with traditional SOA, it performs split in a more refined way, which enhances resource recycling rate and efficiency. The decentralization principle and lightweight communication agreements adopted are more flexible than ESB, which makes possible more accurate service optimization plans after service simplification, improves system maintainability, and shortens product iteration cycle for the upgrading need of the Internet era. For some scholars [15], the cost of MSA service governance and service granularity promotion must keep balance, and distributed development means larger challenges for teams, requiring higher technical costs.

In practical service application, challenges come from both the service and the technique. This section draws a detailed summary of the features of the four architectures and analyzes their key distinctions (Table 1). In terms of hierarchy, monolithic and vertical architectures centralize functional modules of each hierarchy with high coupling degree; SOA uncouples multiple functional modules of vertical and horizontal hierarchies of three or more tiers, but public modules can only be shared on horizontal hierarchies, leading to unthorough uncoupling; the fully self-service flexibility achieved by simultaneous uncoupling on vertical and horizontal hierarchies represents the main characteristic of microservice architecture; however, when putting large projects into practice, development teams cannot comply with all the features and they must consider the integration of irreplaceable systems and promote the flexibility of full uncoupling within acceptable changing rate. Here, the rules of system microservice governance are summed up into two principles:

Principle 1. Regarding the existing service system as services which can participate in microservice architecture, namely, regarding the external systems of each monolithic architecture as independent services clustered by autonomous agents.

Principle 2. Based on the concept of microservices, using SOA architecture, which is controllable for the technical team, to proceed uncoupling upgrade. If the technical routes are closed, then adopt verified upgraded SOA architecture, if not, use the brand new microservice one to explore.

3. Microservice Framework and Key Technologies

Based on the principles of microservice governance, this section designs a whole set of microservice frameworks and common services and describes key functions in microservice development including configuration management, service discovery, circuit breaker, intelligent routing, micro

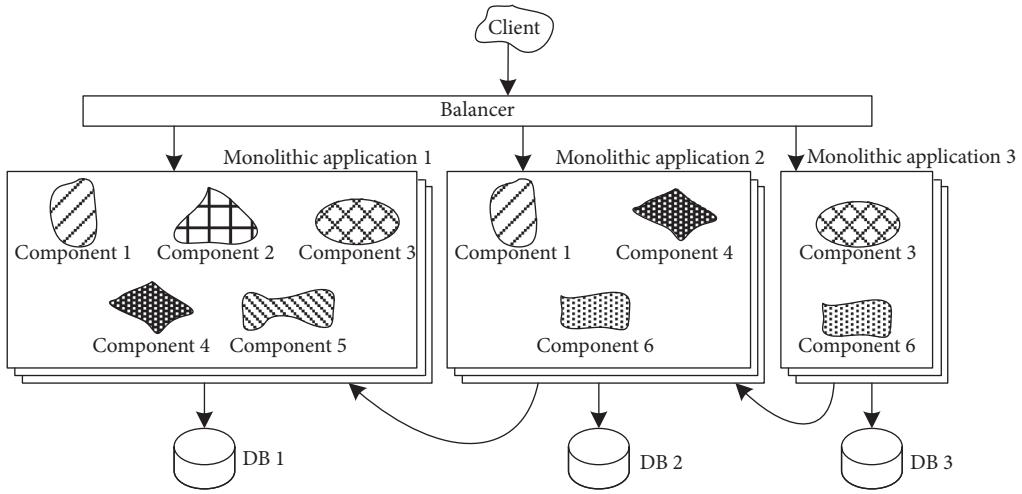


Figure 2: Framework of vertical architecture.

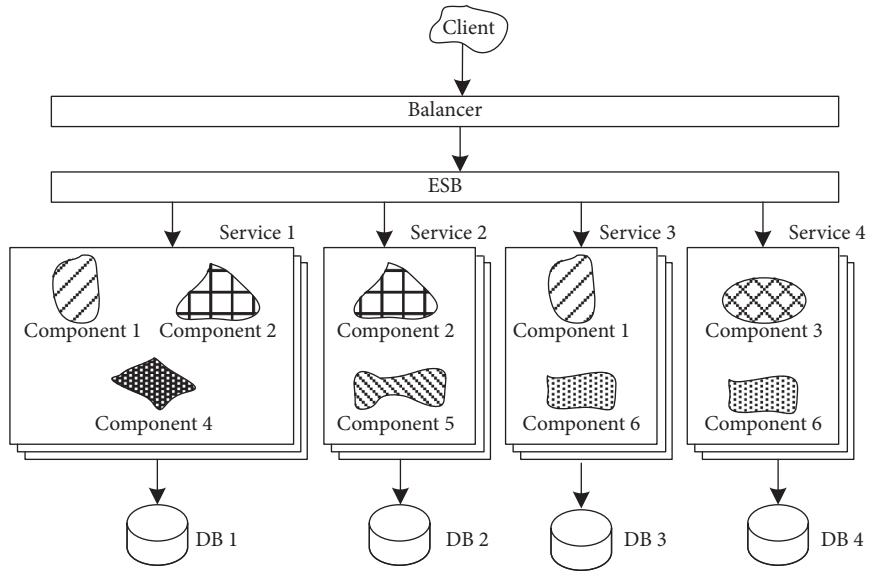


Figure 3: Framework of service-oriented architecture.

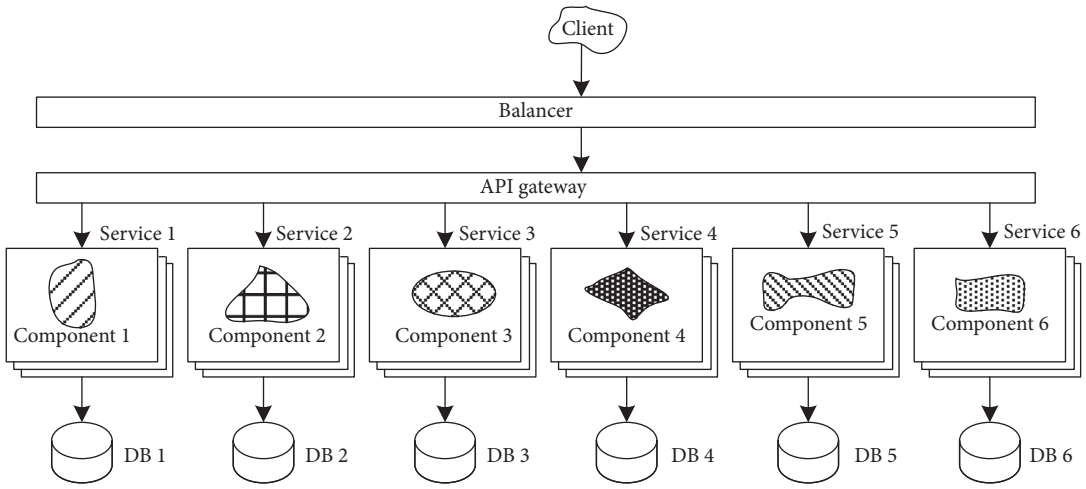


Figure 4: Framework of microservice architecture.

Table 1: Key differences between four architectures.

	Monolithic architecture	Monolithic–vertical architecture	SOA architecture	Microservices architecture
Time	Before 1990s	1990–2000	2000–2010	After 2010s
Feature	Tight coupling	Tight coupling	Light coupling	Decoupled
Advantage	(1) Simple architecture, short development cycle, and low cost (2) It is relatively easy to optimize and improve performance because of less interaction between interfaces	(1) Simple architecture, short development cycle, and low cost (2) Vertical splitting of original monomer projects should not be excessively enlarged (3) Different projects can adopt different technologies	(1) Extraction of common components improves reusability, maintainability, and efficiency (2) Different projects or services can adopt different technologies (3) ESB further reduces the coupling between system interfaces	(1) Finer granularity is more conducive to resource reuse and efficiency improvement (2) Decentralization and lightweight communication protocol (3) The system has strong maintainability and short iteration period
Disadvantage	(1) All functions are integrated in one project, which is not conducive to the coexpansion and maintenance of multiple collaborators (2) Performance optimization depends on clustering, which has bottlenecks in high cost (3) Technology stack constraints	(1) Simple decomposition of monolithic architecture (2) Not conducive to the development and maintenance of large-scale projects (3) Performance expansion has bottlenecks	(1) The boundary between system and service is blurred, which is not conducive to development and maintenance (2) The interface protocols of services are not fixed and there are many kinds of services (3) The extracted service granularity is too large and the coupling between system and service is high	(1) There should be a reasonable balance between the cost of service governance and the refinement of service granularity (2) Challenges to development team and high cost of technology
Scope of application	Small projects	Medium-scale project	Large project	Large projects with frequent and complex interactions

brokers, control bus, distributed sessions, and clustering state management. Also, it analyzes the relationship between the underlying technical principle, logical design, and modules so as to provide guidance and examples on how to rapidly promote microservice governance and realize service and data connection and sharing. Figure 5 shows the concise processing flow diagram of microservice. Key technologies consist of service registration and discovery, remote service calling, circuit-breaker mechanism, service link tracking, and annotation interfaces.

3.1. Service Registration and Discovery. The naming service exists as a basic service in microservice architecture, with name servers as the central node. Each service system defines a service name which is taken as the identifier by the system and registered at the name server. The server identifies each service system by its service name, which provides routing relays for mutual calls between multiple systems. The detailed process is as follows:

Step 1. When the service producer starts, it registers the service it provides at the service registration center

Step 2. When the service consumer starts, it subscribes to the service it needs at the service registration center

Step 3. The registration center returns the address information of the service provider to the consumer

Step 4. The consumer calls the service from the provider

3.2. Remote Service Calling. Remote service calling clarifies the calling protocol for intersystem services, which makes services calling external systems the same as proxy local services. Remote service calling uses the Hypertext Transfer Protocol (HTTP) POST protocol to establish an outgoing channel for data service call. As the HTTP protocol is stateless, JWS authentication method is adopted in system design to ensure secure access to internal resource services and prevent illegal access. Each remote calling must use the token, and the server will verify the token to assess the validity. The basic principles are as follows:

Step 1. The service caller performs remote calling requests according to the service name.

Step 2. The service center picks up the service provider from the service list due to the requested service name.

Step 3. The HTTP connection is established between the service caller and the finder, and data is sent through POST. After receiving the HTTP request, the service provider will first perform validity check and processes the corresponding service according to the request data if the check is passed. When the processing is completed, the result data is returned to the service caller.

3.3. Circuit-Breaker Mechanism. The function of circuit-breaker mechanism is to avoid system failure from spreading. When a system or a function cannot provide

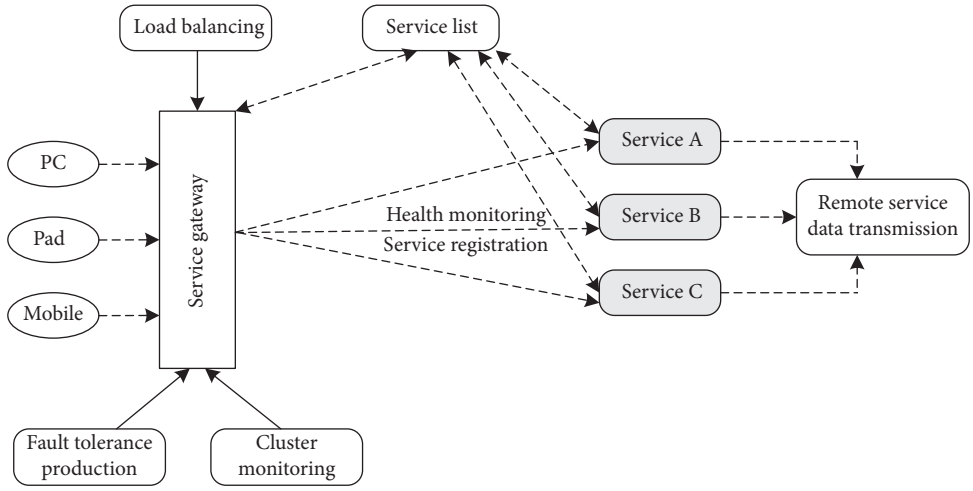


Figure 5: Concise request processing diagram of microservice architecture.

services due to failure, the system should be automatically taken offline from the entire distributed platform. In this way, large-area system crash caused by all requests waiting for a response at that function point can be avoided. The circuit-breaker mechanism is as follows:

Step 1. When a certain threshold is met (the default value is over 20 requests in 10 seconds).

Step 2. When the failure rate reaches a certain level (the default value is over 50% request failure in 10 seconds).

Step 3. When the above threshold is reached, the circuit breaker will turn on.

Step 4. When it turns on, all requests will not be forwarded.

Step 5. After a while (taking 5 seconds as the default value), the breaker is half-open at this moment and will let one of the requests be forwarded. If this succeeds, the circuit breaker will turn off; if not, it will maintain opened and Step 4 will be repeated.

3.4. Service Link Tracking. Service link tracking provides full-link tracking and monitoring during the completion of the entire function, which can clearly show the relationship between each service calling and accurately pinpoint when problems occur. When the system is running, the service link tracking module receives the real-time monitoring data of each microservice system. It mainly includes 4 components:

Component 1: collector, which receives or collects data transmitted by each application.

Component 2: storage, which stores data received or collected in internal storage by default. It currently supports Memory, MySQL, Cassandra, etc.

Component 3: API (query), which is responsible for querying data stored in storage and providing simple data obtained by JSON API, mainly for web UI.

Component 4: Web, which provides simple Web interfaces.

3.5. Annotation Interfaces

Service registration interface @EnableEurekaClient: as an annotation, this interface will go through scanning when the system starts. If the annotation storage is scanned, the configuration information of the naming server is automatically obtained from the configuration file to register the service system.

Remote service call interface @FeignClient: this interface appears as an annotation and the annotation is added when a remote service call is required. After the remote service name is input, the interface obtains the address of the remote service from the naming server and establishes the http connection.

Circuit-breaker interface @EnableHystrix: this interface starts the circuit breaking service, and the corresponding mechanism will start automatically.

Circuit-breaker detection interface @EnableHystrixDashboard: this interface launches the circuit breaking monitoring UI page and visualizes all relevant data on the interface.

Service link tracking interface @EnableZipkinServer: the service link interface starts the health monitoring service, collects trace information from the http protocol from collector. The client terminal calls/api/v1/spans or/api/v2/spans to report the trace information.

4. Microservice Splitting Granularity Decision and Calculation Formula

The core role of microservice architecture is to cope with the growing service capability within the system and the increasingly complex interaction demands between systems. After microservice governance is performed in accordance with the principles and frameworks in the first two sections

of this paper, papers evaluating the effectiveness are quite rare. Obviously, according to the fine-grained service decoupling method in this paper, service updates will be more frequent and the iteration speed will be raised. To ensure that software plays a greater role in overall value creation, the maintainability measurement of microservice system is the most important index.

Definition 1 (maintainability). Referring to the description of Rowe et al. [16], maintainability of microservice system can be defined as how the system can maintain effective and efficient when going through correction, refinement, expansion, or optimization.

Based on Definition 1, the operation of maintainability can be separated into two parts. One is performing correction and refinement when demands are stable; the other is performing expansion and optimization when demands change. Though lots of papers [17] have provided definitions on the maintainability of object-oriented system, there has been no unified conclusion when it comes to microservice system. Thus, on the basis of a large amount of literature research [18–20] and practical experience, this paper discusses from three dimensions related to maintainability, namely, size, coupling degree, and cohesion.

For convenience, we suppose microservice system Y is composed by n services, namely, $Y = \{S_1, S_2, \ldots, S_n\}$, $|Y| = n$. Service S_i has m interaction interfaces, namely, $S_i = \{\mathrm{SI}_1, \mathrm{SI}_2, \ldots, \mathrm{SI}_m\}$, $|S_i| = m$.

Definition 2 (size). The size of a microservice system is the sum size of all its subservices. Obviously, under the same condition, the larger the microservice system is, the lower its maintainability is. The traditional definition of size for a service or operation is related to Lines of Code (LOC) [17]. However, in this paper, the size of the service is represented by the weighted value of the operation amount contained in the exposed interface of a service S, namely, $S_{\mathrm{ws}} = \sum_{\mathrm{SI} \in S} \sum_{O \in \mathrm{SI}} w_l / mk$, in which $W_l (l = 1, 2, \ldots, k)$ is the weighted value of operations within a certain interface (the weighted value can be set according to the amount of parameters and the coarseness of the interface), and the size of microservice system Y is $Y_{\mathrm{ws}} = \sum_{S \in Y} S_{\mathrm{ws}} / |Y|$.

Definition 3 (coupling degree). Coupling degree reflects how much one service depends on others. The lower it is, the higher the maintainability of microservice system is. If service S_i relies on service S_j and vice versa, then this is called interdependence, which is what we have to avoid (and combine the two services into one) in practice. According to [19], S_{IOS} is the importance of a service, showing the amount of consumers depending on service S (the amount of clients of interface SI_i of the called service S). S_{DOS}, the dependence of a service, shows the amount of services depended on by service S (the amount of services of which more than one interface is called by service S). The criticality of a service $S_{\mathrm{COS}} = S_{\mathrm{IOS}} * S_{\mathrm{DOS}}$. Though lower coupling degree means higher maintainability, in the transformation of microservice system services, low coupling degree will constitute a bottleneck. That is because service S will always be called or calls other services and the coupling degree can, to a large extent, help system designers find unreasonable calling.

Definition 4 (cohesion). Cohesion refers to the contribution of the service operation to a certain task or function. There is a positive correlation between it and system maintainability. The connotation of cohesion is quite complex semantically, so it is hard to measure automatically. According to [19], the interface data cohesion (IDC) of service S is defined as the similarity of the parameter type of S's internal interface SI_m, namely, $S_{\mathrm{IDC}} = P_{\mathrm{type}}/m$, in which P_{type} is the data type of interface parameter. When $S_{\mathrm{IDC}} = 1$, the system cohesion is high. Interface usage cohesion (IUC) represents the ratio of the internal interface amount of service S called by the customer and the total amount of interface data of service S, namely, $S_{\mathrm{IUC}} = |\mathrm{SI}_{\mathrm{invok}}|/m$. Similarly, when $S_{\mathrm{IUC}} = 1$, the system cohesion is high. S_{TIC}, the total interface cohesion (TIC) of service S, equals $(S_{\mathrm{IDC}} + S_{\mathrm{IUC}})/2$.

5. Practical Microservice Governance

Shaanxi Medical Products Administration (hereinafter referred to as the Administration) has been using monolithic architecture and upgraded vertical architecture for information construction and connecting with systems of other provincial departments through Web service. By the end of 2016, it has formed a comprehensive service platform covering all aspects of food and drug regulatory system and generalized it to the whole province. However, with the growing functions of the system, the accumulating service data and the provincial-wide user group, the performance of the system has been under great pressure and gradually could not satisfy actual application needs any more. This was manifested in (1) slow response, which makes it failed to return data within a reasonable time and (2) insufficient concurrency support, which means when services are centralized in a certain period, abnormal conditions will happen such as service denial. Besides, as the system became more complex, the cost and risk of new function development, testing, and launch greatly increased. According to the overall requirements of the "Internet + government services" and the 13th Five-Year Plan on Food and Drug Safety of Shaanxi Province and the actual problems existing in the system, experts of the Administration discussed and decided to rearchitect the original system with the new architecture of distributed microservice governance. In this way, the service system can achieve infinite horizontal scaling and gain support in adapting to upgraded demands. The platform also provides the unified data interaction service that truly integrates small systems into a large one.

5.1. Goal of Microservice Governance. According to the requirements of China Food and Drug Administration and the informatization of provincial food and drug administrations, the service framework of the food and drug regulatory system is shown in Figure 6. The system is

mainly composed of three components: the comprehensive service platform, the enterprise service platform, and the collaborative social governance platform. Among them, the first platform is accessed through the unified identity authentication system and serves as the main working platform for supervisors at all levels. After the enterprise service platform is registered and logged in by enterprise users, it can conduct service declaration, process progress inquiry, and receive and give feedback on regulatory information. The third platform requires no registration and login, and one can check or query through the Administration's website or WeChat. It integrates with the newly revised website of the Administration, making data query and utilization more convenient.

Based on the microservice concise processing framework of Figure 5 in Section 3, microservice upgrade and transformation are performed on the original Shaanxi provincial regulatory information system of small-scale food workshops, catering units, and stall keepers. The key tasks and goals include database table management, service splitting of monolithic architecture, workflow customization, and establishment of microservice supporting platforms.

(1) Database table management is to reasonably split some basic information tables that affect performance, making each table as responsive as possible. In terms of the repetitive, unreasonable fields left over from the past, uniform standards are defined. This part mainly contains splitting the enterprise principal information basic table (subtables and the main table are associated through Universally Unique Identifier (UUID) by its uniquely identified subject), reencoding organizational structure (managing the organizational hierarchy by the custom encoding format, and listing administrative region codes as a common attribute in order to flexibly adjust the organization level), splitting the service library (isolate the service library table and sync and share common basic data. Data that belongs to a specific module is regularized, cut from the intersect with other service system data), and replacing triggers with active service calling (use active remote service calling to replace triggers in data upgrade and sync, as the large use of the latter may cause database deadlock)

(2) According to the actual service classification of the Administration, services of the original monolithic system are split into multiple subsystems that can function and be deployed independently, including management of basic user authority, regulatory files, administrative licensing, daily inspection, inspection and handling, random inspection, public notice, and external network report. The relationship of different service systems is teased out. Each service system provides the services they need to release externally so that other systems can call those services remotely. To facilitate management, service systems are not allowed to implement the functions of others. A unified standard for external services is adopted and JSON Web Token (JWT) authentication is used for service calling

(3) In order to improve the flexibility and adaptability of the administrative licensing process, a workflow engine is utilized to realize licensing process customization. More specifically, by defining BussinessKey, enterprise users are linked with process definition which is then connected with the license type through keywords of the predefined license type. The process node is associated with the actual system processing function through predefined keywords

(4) The establishment of microservice supporting system is the core part of microservice transformation. From the perspective of service transformation, each service system provides one or more specific services. In traditional architectures, these systems are totally independent and can hardly utilize specific services of each other. That is to say, many functions are repeatedly established. Through open-source framework, the system of the Administration can achieve basis platform functions such as naming, remote service calling and load balancing, circuit-breaker mechanism, and service self-recovery and service link tracking (Figure 7).

5.2. Results of Microservice Splitting Granularity Decision. According to the microservice framework in Section 3 and the microservice governance goals of the Administration in Section 5.1, service analysis and service-oriented governance have been performed on the original regulatory information system. The system was divided into 12 microservices including administrative licensing, daily supervision, double random inspection, inspection and enforcement, random inspection, risk and credit rating, regulatory archives, data analysis, system maintenance, system monitoring, and information submission. Calculation results of the maintainability indicators of each microservice are shown in Table 2, which shows that microservice transformation greatly raises the scalability and maintainability of the system.

The system of Shaanxi food and drug supervision completed the start-up construction of government procurement procedures in July 2017 and passed the expert acceptance on July 4, 2018. As of June 30, 2018, the system has covered 14 county-level bureaus directly subordinate to the municipal and provincial governments, 104 district and county bureaus, more than 1,400 regulatory agencies, 10,715 supervisors, and 284,544 enterprise users, namely, nearly all food and drug supervisors and food and drug regulatory targets of the province. The supervisory terminal is logged in by 300,000 people per month and the enterprise terminal by 140,000. In total, the system produces 13.64G structural data and 890G unstructured data.

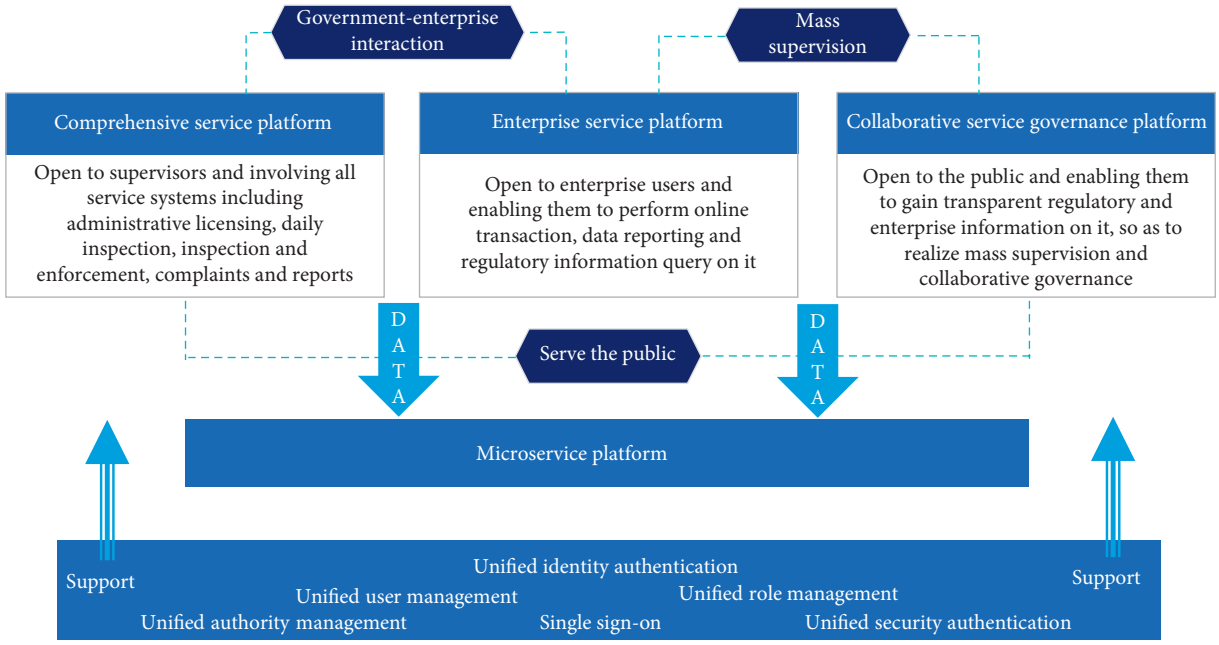


Figure 6: Service framework of the provincial food and drug regulatory system.

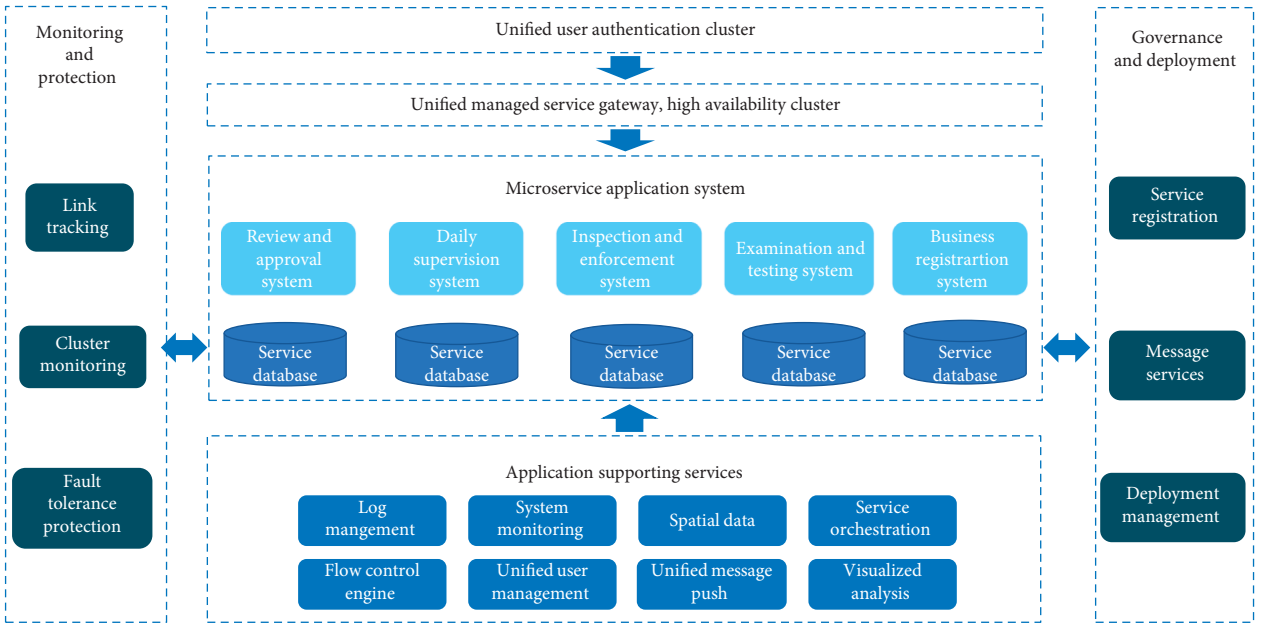


Figure 7: Framework of the microservice-supporting platform.

The comprehensive service system of the Administration implemented the value effect of "three portals, four terminals" (Figure 8). The three portals mean the one-network-based governance for government regulation, the one-network-based transaction for enterprise businesses, and the one-network-based communication for public service, which has achieved the "unified deployment in the province scale and hierarchical application on the prefectural and municipal level" through microservice transformation. The four terminals mean that due to the microservice terminal adaptation function, the system can be accessed through multiple channels including PC, Android, and IOS mobile terminals and WeChat official accounts. The microservice transformation result of Shaanxi province has won a large round of applause from acceptance review experts and insiders and appeared in the special reports in the "Annual Review during the Two Sessions" program of CCTV and Office of the Central Cyberspace Affairs Commission. As one of the "2018 National Intelligent Regulation Exemplary Cases," it has been generalized and replicated in 11 provinces and cities.

Table 2: Calculation results of the maintainability indicators.

No.	Service name	Interface number	Operation number	S_{ws}	S_{IOS}	S_{DOS}	S_{COS}	S_{IDC}	S_{IUC}	S_{TIC}
S_1	Administrative licensing (API)	5	8	0.35	8	3	24	0.9	0.60	0.75
S_2	Daily supervision (XML)	3	5	0.54	5	5	25	0.9	1.67	1.28
S_3	Double random check (API)	3	3	0.67	3	5	15	0.8	1.67	1.23
S_4	Enforcement of inspection (WSDL)	4	5	0.45	2	5	10	1	1.25	1.13
S_5	Sampling inspection (JSON)	2	5	0.70	2	3	6	0.7	1.50	1.10
S_6	Risk rating (API)	3	3	0.67	2	3	6	1	1.00	1.00
S_7	Credit rating (API)	3	3	0.67	2	3	6	1	1.00	1.00
S_8	Regulatory archives (API)	5	2	0.70	10	2	20	1	0.40	0.70
S_9	Data analysis (API)	2	10	0.60	2	10	20	0.3	5.00	2.65
S_{10}	System maintenance (XML)	3	5	0.53	2	10	20	1	3.33	2.17
S_{11}	System monitoring (XML)	2	10	0.60	1	10	10	1	5.00	3.00
S_{12}	Information submission (WSDL)	2	2	1.00	1	8	8	1	4.00	2.50

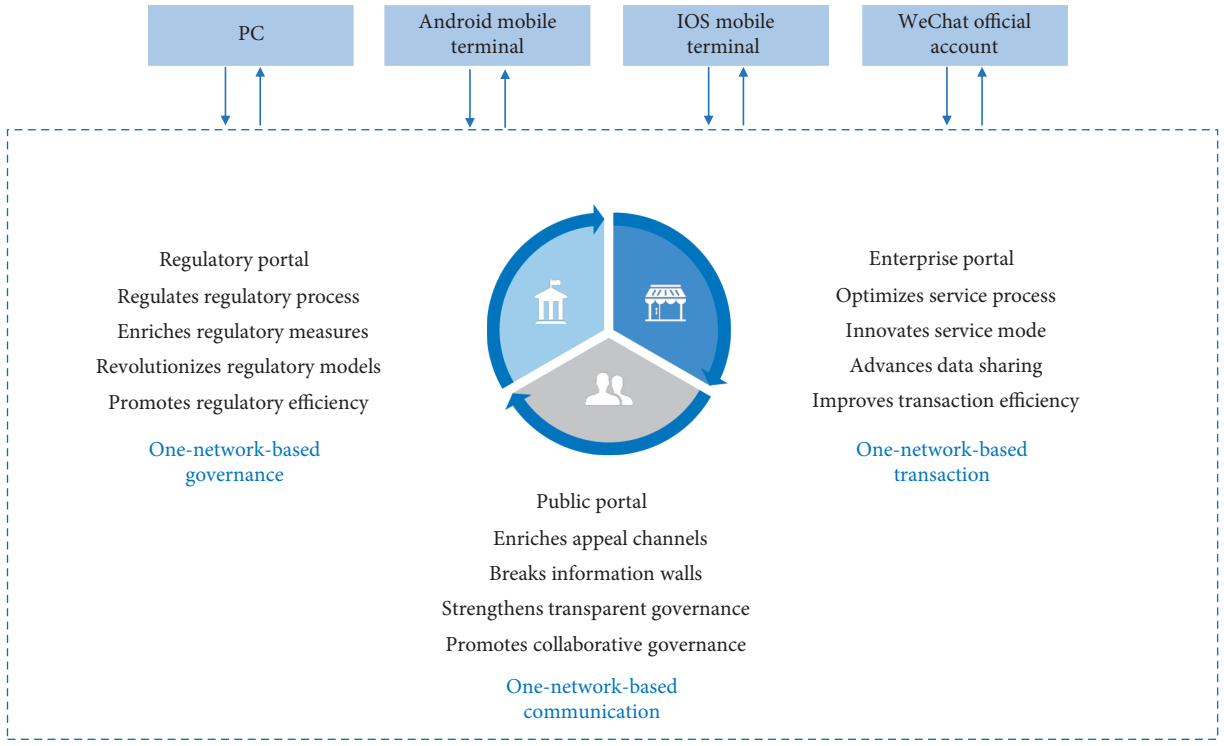


Figure 8: Value effect of "three portals, four terminals."

6. Conclusion and Prospect

The data-driven concept can innovate the comprehensive social governance mechanism effectively, and the premise of utilizing data-intensive research paradigm is the full sharing and integration of data. This paper first summarizes the critical meanings of government data sharing, introduces the four architectures of data integrated governance in a systematic way, and analyzes their differences, advantages, and disadvantages. Then, it provides the basic framework and the maintainability measure indexes of microservice from the perspective of practical application. In the end, it takes the microservice governance of a provincial food and drug regulatory system as an example to verify the framework and measure indexes proposed, which, as the practical result shows, promote data sharing and the expandability and maintainability of the system. Through microservice governance, collaborative social governance featuring "one-network-based governance, transaction, and communication" can be achieved. It is an exemplary mode worth popularization.

Microservice architecture is a practical technology derived from applications, most of which are based on individual experience or understanding for granularity segmentation of microservices. This paper firstly analyzed and discussed the split decision-making process of microservice from a rigorous theoretical analysis perspective, which has certain innovation significance in this field. However, the theoretical decision-making in this paper is only from a simple application and maintenance perspective, without considering some other factors. Therefore, more factors need to be integrated for rigorous theoretical analysis and cases in the subsequent research.

Data Availability

The data used to test the matching decision model of this study are included within the article. Additional data can be provided by the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This research was supported by the Shaanxi Provincial Department of Education serving the Special Project of Local Enterprises under Grant no. 2019TJ034; Research Project of Key Projects of Statistical Science in Shaanxi Province under Grant no. 19JC015; Xi'an Science and Technology Plan University Talent Service Enterprise Project under Grant no. GXYD7.7; and Xi'an Polytechnic University Doctoral Research Start-Up Fund under Grant no. 107020342.

References

[1] C. Kucklick, *The Granular Society*, CITIC Group Press, Beijing, China, 2017.

[2] C.-h. Wang, "Evolution and innovation of "comprehensive governance" in China," *Journal of Beijing Administrative College*, vol. 2, pp. 42–46, 2015.

[3] C.-H. Chen, "A cell probe-based method for vehicle speed estimation," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E103.A, no. 1, pp. 265–267, 2020.

[4] Z.-m. Zeng and Y. Qian-wen, "Research on the data curation system of urban public security oriented to the fourth paradigm," *Information Studies: Theory & Application*, vol. 2, pp. 82–87, 2018.

[5] Ye Tian, Bo Yuan, and Li Ting-li, "A massive and heterogeneous data storage and sharing strategy for Internet of things," *Acta Electronica Sinica*, vol. 44, no. 2, pp. 247–257, 2016.

[6] L. I. Xiao-tao, H. U. Xiao-hui, G. U. O. Xiao-li, and W.-n. LU, "Complicated information sharing technology based on metadata," *Systems Engineering and Electronics*, vol. 37, no. 3, pp. 700–706, 2015.

[7] L. Ming, "Large data technology and public security information sharing ability," *E-government.*vol. 6, pp. 10–19, 2014.

[8] B. Lake, "*An empirical evaluation of an agile modular software development approach: a case study with ericsson*," M.S. thesis, Stockholm University, Stockholm, Sweden, 2012.

[9] R. Arnon, *SOA Patterns*, Manning Shelter Island, Shelter Island, NY, USA, 2012.

[10] D. Sprot and L. Wilkes, *Understanding Service Oriented Architecture*, Microsoft Developer Network, 2004, https://msdn.microsoft.com/en-us/library/aa480021.aspx#aj1soa_topic5.

[11] J. Thönes, Micro-services, IEEE Software, ISSN 0740-7459/15, 2015.

[12] C.-H. Chen, F. Song, F.-J. Hwang, and L. Wu, "A probability density function generator based on neural networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 541, Article ID 123344, 2020.

[13] G. Gruman and A. Morrison, "Micro-services: the resurgence of SOA principles and an alternative to the monolith," *Technology Forecast: Rethinking Integration*, vol. 1, 2014.

[14] M. Fowler, "Micro-services," 2014, http://martinfowler.com/articles/micro-services.html.

[15] M. Vianden, H. Lichter, and A. Steffens, "Experience on a micro-service-based reference architecture for measurement systems," in *Proceedings of the 21st Asia-Pacific Software Engineering Conference*, IEEE, Jeju, South Korea, 2014.

[16] D. Rowe, J. Leaney, and D. Lowe, "Defining systems architecture evolvability-a taxonomy of change," in *Proceedings of the International Conference and Workshop: Engineering of Computer-Based Systems*, pp. 45–52, Jerusalem, Israel, December 1998.

[17] M. Perepletchikov, C. Ryan, and F. Keith, "Comparing the impact of service-oriented and object-oriented paradigms on the structural properties of software," *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 3762 LNCS*, pp. 431–441, Springer, Berlin, Germany, 2005.

[18] C.-H. Chen, F.-J. Hwang, and H.-Y. Kung, "Travel time prediction system based on data clustering for waste collection vehicles," *IEICE Transactions on Information and Systems*, vol. E102-D, no. 7, pp. 1374–1383, 2019.

[19] M. Perepletchikov, C. Ryan, and F. Keith, "Cohesion metrics for predicting maintainability of service-oriented software," in *Proceedings of the Seventh International Conference on Quality Software (QSIC 2007)*, IEEE, Portland, OR, USA, pp. 328–335, 2007.

[20] D. Rud, A. Schmietendorf, and R. R. Dumke, "Product metrics for service-oriented infrastructures," in *Proceedings of the IWSM/MetriKon*, Montreal, Canada, 2006.

Hindawi
Discrete Dynamics in Nature and Society
Volume 2020, Article ID 4827503, 21 pages
https://doi.org/10.1155/2020/4827503

Research Article

Multiattribute Supply and Demand Matching Decision Model for Online-Listed Rental Housing: An Empirical Study Based on Shanghai

Lingyan Li,[1] **Jiangying An**,[1] **Yan Li**,[2] **and Xiaotong Guo**[1]

[1]*School of Management, Xi'an University of Architecture and Technology, Xi'an 710055, China*
[2]*School of Management, Xi'an Polytechnic University, Xi'an 710048, China*

Correspondence should be addressed to Yan Li; sayidli@xpu.edu.cn

Received 22 April 2020; Revised 10 June 2020; Accepted 23 June 2020; Published 1 August 2020

Guest Editor: Chi-Hua Chen

The mismatch between the supply and demand of online-listed rental housing (ORH) is an important factor restricting the operational efficiency of online rental service platforms. However, extant literature pays little attention to this problem. This study proposes an ORH multiattribute supply and demand matching decision model based on the perceived utility of matching both sides of this market. The model considers the multiattribute information of ORH, such as area, transportation, rent, room, and interior decoration, and quantifies their perceived utility values based on the theory of disappointment. Thereafter, we construct the matching decision model and verify it for feasibility by applying it to Shanghai's ORH supply and demand information—our empirical case. The results show that this method can be applied to online rental housing platforms and meet the supply and demand matching requirements to the greatest extent. The constructed model takes into account the perceptions of both supply and demand parties, may promote the effective matching of ORH supply and demand, and bears theoretical implications for the improvement of rental housing matching in ORH platforms.

1. Introduction

Matching supply and demand (SAD) directly affects the success of online trading. However, regarding rental housing, efficiency in matching SAD is not high [1]; many countries around the world have rental housing markets characterized by mismatches in SAD [2]. In terms of rental expenses, the US' traditional housing market, for example, offers properties with higher rental prices than residents can afford [3]. Other examples are Germany, where rented homes do not match regional SAD [4], and Nigeria, where the speed and scale of housing supply are smaller than the demand, causing inefficiency in the country's ability to meet the housing needs of low-income earners [5]. Similarly, in China, the rental housing SAD structure is incongruent [6, 7]. This shows that traditional rental housing supply structures are incapable of matching personalized demands, and the mismatch between the quality and quantity of ORH leads to this discrepancy between the rental housing market's SAD, which seriously hinders the efficiency of rental housing service platforms. Therefore, not only can the creation of an effective matching method for online rental housing platforms solve the operational inefficiency of the rental housing market but can also promote the rapid development of the rental market.

To promote the effective matching of rental housing online, some scholars have made various attempts to study the SAD matching of ORH. So far, the matching model of ORH is mainly a balanced search model, whose purpose is to reduce search costs [8, 9]. There are a few models to measure the quality of supply and demand matching based on the similarities between SAD [10]. There are also models to establish a double-sided matching theory to solve the problem of public rental housing [11]. At the same time, some scholars pointed out that SAD patterns consider various housing attributes in the house matching selection

process [12, 13] and that these attributes affect ORH SAD matching [14]. For example, price floors play a central role between SAD [15], while issues around house matching and comfort, transportation, the surrounding environment, and landlord services are specific indicators of tenant concerns [16]. Location is also a factor of housing mismatch [2]. The size and location of a house are considered to be necessary conditions for appropriate house matching [14]. Therefore, housing attributes play a very important role in balancing the SAD of ORH.

Although the aforementioned models proposed by extant studies can reduce the cost of searching in the matching process—considering that housing multiattribute indicators are factors in the SAD matching process—they do not promote the effective matching of ORH or the improvement of matching satisfaction between the SAD sides. In fact, when a matching decision occurs in the actual process, most subjects are in limited rationality [17]; their expected values are limited by existing information and their own expectations [18, 19]. Thus, the final matching pair may have a higher or lower value than the expected value, leading to below-optimal satisfaction with the matching results. In addition, perceived utility is a product of a demanders' assessment of the services, values, and functions of the required items compared to their own expectations, how they weigh the perceived benefits and costs, and the value they finally put on the required items [20]. In assessing profit and loss values, the theory of disappointment reflects the same difference between actual and expected results [21]. Loomes and Sugden also believed that "disappointment" and "excitement" are key factors in making rational choices [22]. Therefore, using disappointment theory to quantify the perceived utility of both SAD can reflect the psychological activities of both sides.

In recent years, with the acceleration of urbanization, the housing problem of China's floating population has become increasingly prominent. In 2018, the floating population reached 241 million, of which 210 million required rental housing, covering an area of 6.73 billion square meters. In addition, large and medium-sized cities in China are required to build a rental housing platform to promote the matching efficiency of rental housing. However, the housing rental market is undeveloped and has many problems, such as low supply efficiency, lack of diverse supply systems, and low capability to meeting people's housing demand [6, 23]. Among them is Shanghai, the central city of China, with a relatively large annual flow of floating population and a great demand for rental housing. Therefore, taking Shanghai as our case study may help understand and reflect the problems of other cities. This is of great significance for promoting the development of the rental housing market.

In online rental housing transactions, the real estate network platform involved is a two-sided market that connects the SAD sides of rental housing through which various agents interact with one another, and the transactions cater to the needs of all the users (both SAD). Therefore, this study attempts to apply the two-sided matching theory to the housing rental market. To improve the satisfaction of both SAD sides, this study uses disappointment theory to construct quantitative perceived utility value matrices. First, we start by quantifying the utility of housing as perceived by both SAD sides based on the multiattribute matching characteristics of both sides. Second, we construct a multiattribute SAD matching decision model for ORH platforms. Finally, we adopt online SAD information from Shanghai, China, to ensure that the multiattribute information of rental housing has been identified and to empirically test the validity and rationality of the model, safeguarding the construction of an innovative method for ORH SAD matching.

The contributions of this study are twofold. First, the study constructs an effective SAD matching model for online rental housing platforms and enriches the application of the two-sided matching theory. Currently, research on two-sided matching is mostly applied in the marriage market, labor market, schools, and hospitals, but is rarely used in the rental housing market, especially for application in online rental housing service platforms. Therefore, this study undertakes to verify the applicability and effectiveness of the two-sided matching theory in the rental housing market. This provides a reference and a theoretical method for cracking the SAD matching problem that has plagued online rental housing platforms with multisubject and multilevel supply methods. Second, the attributes of ORH are determined and measured. These can be easily extended to other cities and online platforms, providing reference to the feasibility of analyzing multiattribute decision-making matching problems. This also helps to improve the overall satisfaction and quality of SAD matching efforts by considering the expectations and real values of both SAD sides. In sum, this study is valuable for improving the efficiency of SAD matching for online platforms, as well as the sustainable development of the rental housing market.

The remainder of this article is arranged as follows: Section 2 presents a detailed literature review. Section 3 introduces the research materials and methods. The empirical results and discussion are presented in Section 4. Finally, Section 5 summarizes the conclusions and implications of this study.

2. Literature Review

2.1. Disappointment Theory. The disappointment theory, proposed by Bell [21] and Loomes and Sugden [22], is based on psychological reactions caused by the comparison of actual results with expectations [24]. Fundamentally, the disappointment theory suggests that when actual results are better than expected, policymakers will be ecstatic, when actual results are worse than expected, manufacturers will be disappointed, and that the greater the disparity between the outcome and the expectation, the greater the disappointment [21]. So far, there are three main contemporary disappointment models [25, 26]. The first is the *Bell–Loomes and Sugden disappointment model*. This model essentially quantifies the "modified expected utility" of human satisfaction in the case of disappointment. The second is Gul's *disappointment aversion* (DA) model, which decomposes the participants' psychological cognition into disappointment

and elation parts with respect to a new certainty equivalent. The third and the last model is Cillo and Delquié's *disappointment without prior expectation* model. This model does not assume that the decision maker sets a single reference point of comparison prior to a certain experience. Taking this theory into account and the idea that actual utility is a combination of subjective utility functions and disappointment and exhilaration functions [21], this paper uses the disappointment theory to calculate the perceived utility values of both SAD sides.

In recent years, many scholars have quantified the satisfaction of matching objects based on the disappointment theory. For example, Quan et al.'s study explored customer psychological satisfaction based on customer disappointment and pleasure behaviors, and their method promotes retailer pricing and inventory decisions [27]. Similarly, Ma et al. established a SAD matching model for the electricity retail industry by considering the disappointment of sensitive customers and new electricity retail companies by considering the pleasant psychological perception and obtaining the perceived utility of both parties. This method is verified to be beneficial for the management of a high-quality electricity market. Zhang et al. built a matching optimization model that maximizes satisfaction by calculating the disappointment and excitement associated with both SAD sides based on the disappointment theory [28]. Zhao et al. proposed a two-sided matching model to elaborate the preference order of SAD by considering the participants' psychological perception. Using the functions of disappointment and cheerfulness, this model effectively improves the carpool matching problem [26]. Another example is that by Fan et al., who proposed a two-way matching method that considers the psychological behavior of the agents of both parties and constructed a two-objective optimization model to obtain satisfactory matching results [29]. These models maximize the satisfaction of both SAD sides, and by considering the degree of their disappointment and elation, these models effectively reflect the perception activities of both sides, which is conducive to promoting the improvement of the satisfaction of SAD.

2.2. Two-Sided Matching. The two-sided matching theory originated from a study on marriage matching that focused on satisfying both parties with matching results according to the requirements of SAD sides [30]. Gale and Shapley, from which Shapley became the 2012 Nobel laureate in economics, first proposed the G-S algorithm for stable matching based on the strong preference information given by both men and women [31]. Later, Roth, another Nobel laureate in economics, applied the G-S algorithm to the hospital internship market [30]. Since then, the two-sided matching theory has been developed and widely used in various fields such as staff and job matching, data trading market matching, commodity trading issues, and college admission issues. At the same time, some scholars have proposed a variety of matching methods, including the deferred algorithm [32], evolutionary algorithm [33, 34], and multiattribute preference decision-making methods [35]. Nonetheless, the two-sided matching theory has proved to have a wide range of practical applications.

With the popularity of the Internet, some scholars have begun to talk about the application of the two-sided matching theory to the SAD matching problems of many online transactions. For the second-hand housing market, taking the online rental-sale matching problem as an example, Wang et al. established a two-sided matching decision method based on heterogeneous information and association, the feasibility of which was verified [36]. In the electronic market environment, Gao et al. provided a stable multilateral automated negotiation system model and showed that this model can promote effective matching between the SAD sides. For the power purchase market [37], Kong et al. proposed a SAD matching model for power retailers and buyers of cellular networks based on the two-sided matching theory, which solves the energy management problem of microgrid-connected cellular networks for the small- and medium-sized market of the sharing economy [38]. Xia et al. constructed a two-sided matching model that promotes the maximization of the interests of buyers, sellers, and agents and improves the matching efficiency thereof. For the hospital's expert-outpatient matching appointment system [39], Yang et al. constructed a matching model based on the two-sided matching theory that improves the effectiveness of medical treatment [40]. These models built by scholars based on the two-sided theory have solved the problem of matching SAD in different fields.

Similarly, in the existing literature, the two-sided matching theory has been widely applied to online trading platforms and has improved the matching efficiency between SAD. Nevertheless, its application in ORH transactions is limited. In addition, existing two-sided matching decision-making methods rarely consider the psychological perception of disappointment or elation when the subjective desires of both parties match and when the psychological perceptions are closely related to the satisfaction of both parties in the final matching scheme. The satisfaction of the two parties needs to effectively characterize the psychological perception of both parties. Moreover, in the process of matching ORH SAD, both sides will consider the multiattribute information of housing [12], but there are very few studies covering the multiattribute two-sided matching model. Therefore, this paper proposes a multiattribute supply-demand matching method for online rental listings based on the perceived utility.

On online house rental platforms, tenants and landlords can be regarded as two unconnected sets. Each set will evaluate the actual value of the property according to its own psychological expectations. The main goal of a matching method in this case is to promote effective matching between SAD based on the tenants' and landlords' mental perception. Therefore, this study aims to apply the two-sided matching theory to the SAD matching problem of online rental housing services to accurately characterize the SAD perceptions of both sides and to quantify the perceived utilities of both SAD based on the disappointment theory to promote satisfaction over the matching results.

3. Materials and Methods

3.1. Framework and Schematic Diagram of the Multiattribute SAD Matching Decision Model of ORH

3.1.1. Description and Assumption. On online rental platforms, three parties are involved: suppliers, demanders, and independent intermediaries or brokers. Suppliers on online rental platforms are usually landlords with vacant houses, the demanders are the potential tenants (those with housing needs), while independent intermediaries or brokers are service organizations that assist in matching the SAD sides based on the information provided by the suppliers and demanders. In the matching process, the brokers satisfy the requirements of the SAD parties to the greatest extent according to the information provided by them. In this study, the interactions of the suppliers and demanders can be explained with a SAD matching framework as done in the following. All variable assumptions and descriptions in the paper are shown in Table 1.

This study describes the matching characteristics of SAD sides. If the tenant is T, then $T = \{T_1, T_2, \ldots, T_m\}$, where T_i is the i tenant, and $i \in M = \{1, 2, \ldots, m\}$. If a collection of landlords is L, then $L = \{L_1, L_2, \ldots, L_n\}$, where L_j is the j landlord, and $j \in N = \{1, 2, \ldots, n\}$, and $m \leq n$. If a collection of online rental house listings displayed by the landlords on the online real estate platform is H, then $H = \{H_1, H_2, \ldots, H_n\}$, where H_j is the j online-listed rental house.

If a collection of housing attributes for evaluating ORH is C, then $C = \{C_1, C_2, \ldots, C_k\}$, where C_l is the l attribute of ORH, $l \in K = \{1, 2, \ldots, k\}$. The factors for evaluating ORH are region, rent, transportation convenience, units of houses, decoration, house supporting facilities, online ratings, and so on. These attributes are divided into hard requirements and soft requirements. Hard requirements are equitation requirements, while soft requirements are divided into benefit-orientated requirements and cost-orientated requirements. The higher the attributes, the better the benefit-orientated requirements; the lower the attributes, the better the cost-orientated requirements. The tenants determine the expected value and weights of the specific attributes, while the landlord determines the evaluation value and attribute weights of the specific attributes such as their actual needs. Among them, $E = [e_{il}]_{m\times k}$ is the expectation level matrix of tenant T_i regarding the ORH, and e_{il} is the expectation level of tenant T_i regarding the attributes of ORH C_l. $R = [r_{jl}]_{n\times k}$ is the real evaluation matrix of landlord L_j regarding the ORH, and r_{jl} is the real evaluation value of landlord L_j regarding the attributes of ORH C_l. Then, tenants give the attribute weight vector $w_{il} = \{w_{1l}, w_{2l}, \ldots, w_{il}\}$, where w_{il} represents the tenant's T_i description of the importance of the attribute C_l, $\sum_{l=1}^{k} w_{il} = 1$, $0 \leq w_{il} \leq 1$, where w_{jl} is the landlord's L_j description of the importance of the attribute C_l, $\sum_{l=1}^{k} w_{jl} = 1$, $0 \leq w_{jl} \leq 1$. Then, according to the attribute values given by tenants and landlords, the weights of each attribute are determined by the entropy weight method.

Based on the above analysis and definition, the multiattribute SAD matching decision model for ORH can be described in Figure 1. The first step is the landlords submitting the real multiattribute evaluation value of the houses to the broker and the broker displaying the multiattribute information of those houses on the online rental platform. The second step is the tenants searching for the desired rental house on the online rental housing platform according to the expected value of each attribute of the rental houses. The third step involves the tenants measuring the real attribute value of the house through the left expected attribute value $C_1, C_2, C_3, \ldots, C_p$ and the landlords measuring the expected attribute value of the tenants through the right expected attribute value $C_1, C_2, C_3, \ldots, C_k$.

3.1.2. Multiattribute Supply-Demand Matching Decision Model Development. The schematic overview of the whole paper, as shown in Figure 2, presents processes and associated techniques used in this study for the construction of the multiattribute SAD matching decision model for ORH based on the perceived utility. The major steps of this process include (a) selecting and measuring housing multiattribute indicators for landlords and tenants to evaluate rental housing, respectively, (b) calculating the profit and loss matrices for each attribute of the landlords and tenants, (c) transforming the landlords' and tenants' profit and loss matrices into normalized profit and loss matrices, (d) calculating the perceived utility matrix for each attribute of the landlords and tenants, (e) determining the weight of each attribute, (f) transforming the landlords' and tenants' perceived utility matrices of each attribute into a comprehensive perceived utility matrix, (g) establishing a multiattribute SAD matching decision model for ORH, and (h) verifying the feasibility of the model based on the actual case.

From Figure 2, the schematic overview of the multiattributes based on the perceived utility is established for supplier-demander interactions on online rental platforms. Through the web crawler method, the expected values and evaluation values of various housing attributes for both SAD were collected. This was followed by selecting seven multiattribute indicators of housing through word frequency analysis and semantic network analysis. These include district, traffic convenience, rent, living room, house supporting facilities, decoration, and online ratings. These attributes were then measured. Moreover, the profit and loss matrices of each housing attribute of the SAD were obtained based on the profit and loss function method. The method can measure the benefits and costs of both SAD. Next, we transformed the profit and loss matrices of the SAD into standardized profit and loss matrices by normalizing functions, which can unify the dimensions of each housing attribute. Furthermore, the perceived utility matrix of each housing attribute of the SAD was obtained based on the disappointment and elation function. Considering the interaction between the housing attributes, the entropy weight method was used to obtain the weight of each housing attribute, and the perceived utility matrix of both the SAD was converted into a comprehensive perceived utility matrix. In addition, based on a comprehensive perceptual utility matrix, the multiattribute SAD matching decision model for ORH was proposed. The model was then converted into a single-objective optimization model by a linear weighted method. Finally, to illustrate its

Table 1: Variable descriptions.

Nomenclature	
T	Collection of tenants
L	Collection of landlords
H	Collection of rental housing
C	Collection of rental housing attributes
E	The expectation level matrix of the tenant
e_{il}	Level of expectations given by the i tenant for the l attribute of online rental houses
R	The real evaluation matrix of the landlord
r_{jl}	Level of real evaluation given by the j landlord for the l attribute of online rental houses
w_{il}	The weight given by the i tenant to the l attribute of ORH
w_{jl}	The weight given by the j landlord to the l attribute of ORH
D_l	Profit and loss matrix of the tenant under the l attribute
d_{ijl}	Profit and loss value of the i tenant with respect to the j landlord under the l attribute
p_{ij}	Transaction rents
F_l	Profit and loss matrix of the landlord under the l attribute
a_{ijl}	Profit and loss value of the j landlord with the i tenant under the l attribute
D_l'	Normative profit and loss matrix of the i tenant under the l attribute
d_{ijl}'	Normative profit and loss value of the i tenant with respect to the j landlord under the l attribute
F_l'	Normative profit and loss matrix of the j landlord under the l attribute
a_{ijl}'	Normative profit and loss value of the j landlord with the i tenant under the l attribute
V_l^b	Tenant's perceived utility matrix under attribute l
V_l^s	Landlord's perceived utility matrix under attribute l
V	Tenant's comprehensive perceived utility matrix
V'	Landlord's comprehensive perceived utility matrix
x_{ij}	0-1 variable
Z	The overall satisfaction value
Greek symbols	
α	The delight parameter
β	The disappointment avoidance parameter
Abbreviations	
SAD	Supply and demand
ORH	Online-listed rental housing

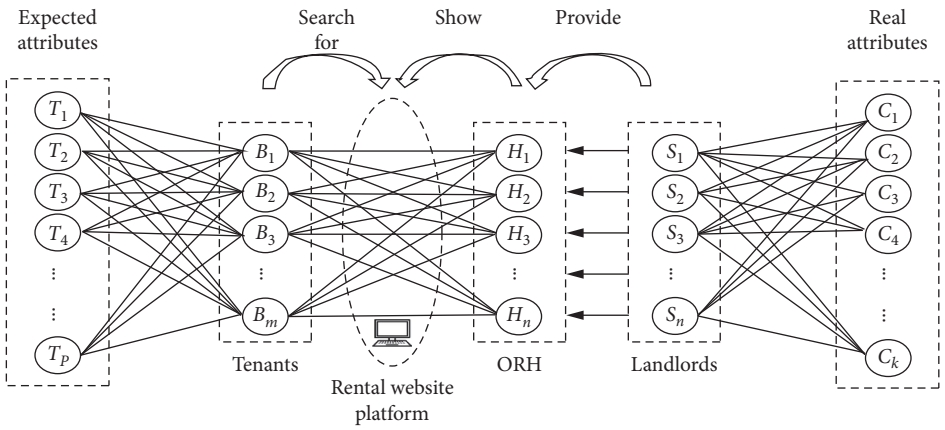


Figure 1: Multiattribute SAD matching decision model for the ORH matching process.

applicability and test its effectiveness, the model was applied to evidence from Shanghai.

3.2. Multiattribute SAD Matching Decision Model for ORH Platform Transactions Based on the Perceived Utility

3.2.1. Basis for the Construction of Profit and Loss Matrices. For both SAD sides, benefits and losses occur during the process of matching the SAD of ORH. These benefits and losses are subject to psychological perception, and psychological perception is closely related to the satisfaction of both SAD sides in the final matching scheme. This study specifically describes the values of the profits and losses of both suppliers and the demanders as follows.

When the expectation value of tenant T_i for an attribute of the rental house C_l is inferior to the real evaluation value of landlord L_j in terms of this attribute, profits occur;

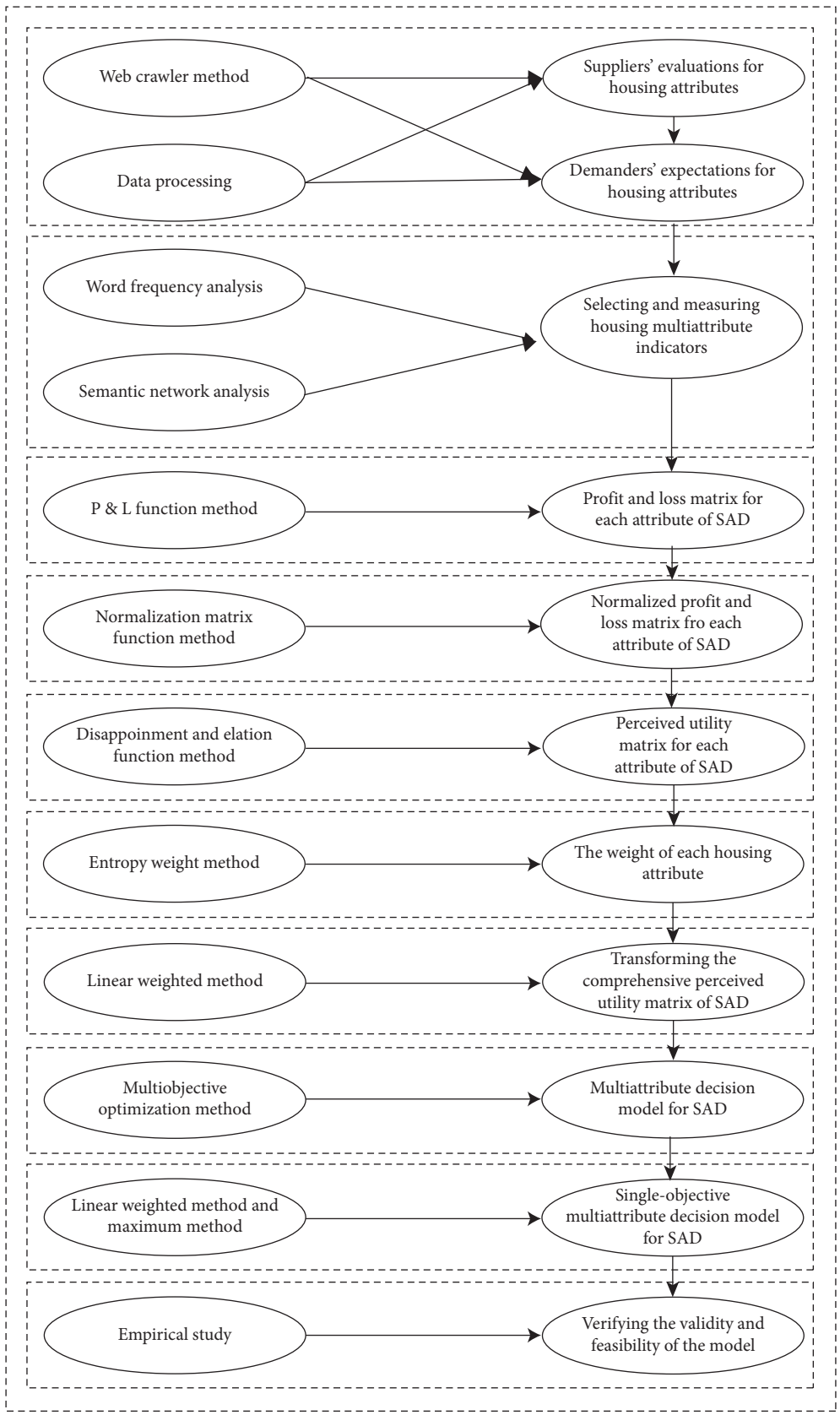


Figure 2: Schematic overview of the multiattribute SAD matching decision model for ORH based on the perceived utility.

otherwise, losses occur. The above benefits and losses are quantified as follows.

If $D_l = [d_{ijl}]_{m\times n}$ is the profit and loss matrix of tenant T_i under attribute C_l, then d_{ijl} is the profit or loss of landlord L_j under attribute C_l. Considering that tenant T_i will put forward different requirements according to their own expectations, d_{ijl} is calculated as follows.

When attribute C_l is rent, in order to ensure the fairness of online transactions, the transaction rent is set to the middle value of the tenants'. expected rent and the landlords' L_j expected rent, and the calculation equation of the transaction rents p_{ij} and d_{ijl} are as follows:

$$p_{ij} = \frac{r_{jl} + e_{il}}{2}, \quad i \in M,\ j \in N,\ l \in K. \tag{1}$$

When attribute C_l is rent, d_{ijl} is calculated by the following three cases [41]:

Case 1: when $e_{il} = r_{jl}$, the tenant's T_i performance to landlord L_j is neither a gain nor a loss.

Case 2: when $e_{il} < p_{ij} < r_{jl}$, the tenant's T_i performance to landlord L_j is a loss. The loss would be $1 - (p_{ij}/e_{il})$.

Case 3: when $e_{il} > r_{jl}$, the tenant's T_i performance to landlord L_j is gain. The gain is 1.

Therefore, the equation for calculating d_{ijl} is as follows:

$$d_{ijl} = \begin{cases} 0, & e_{il} = r_{jl}, \\ 1 - \dfrac{p_{ij}}{e_{il}}, & e_{il} < p_{ij} < r_{jl}, \\ 1, & e_{il} > p_{ij} > r_{jl}, \end{cases} \quad i \in M, j \in N, l \in K. \tag{2}$$

When attribute C_l is not rent, there are three forms of requirements, Wang et al. gave everyone's gains and losses based on expectations, and the equation is as follows [42]:

(1) When tenant T_i has a hard requirement of attribute C_l, hard constraints refer to covenants that must be satisfied by bundle conditions. When the value of e_{il} is equal to the value of r_{jl}, tenant i and landlord J can be matched. When the value of e_{il} is not equal to the value of r_{jl}, tenant T_i and landlord L_j cannot be matched. Then, the equation for calculating d_{ijl} is as follows:

$$d_{ijl} = \begin{cases} 1, & r_{jl} = e_{il}, \\ 0, & r_{jl} \neq e_{il}, \end{cases} \quad i \in M,\ j \in N,\ l \in K. \tag{3}$$

(2) When tenant T_i has a benefit requirement of attribute C_l, the larger the attribute value of the benefit-type attribute, the better. d_{ijl} is calculated by the following three cases:

Case 1: when $e_{il} = r_{jl}$, the tenant's T_i performance to landlord L_j is neither a gain nor a loss under attribute C_l.

Case 2: when $e_{il} < r_{jl}$, the tenant's T_i performance to landlord L_j is a gain under attribute C_l. The gain is $r_{jl} - e_{il}$.

Case 3: when $e_{il} > r_{jl}$, the tenant's T_i performance to landlord L_j is a loss under attribute C_l. The loss is $r_{jl} - e_{il}$.

Therefore, when attribute C_l is a benefit requirement, the equation for calculating d_{ijl} is as follows:

$$d_{ijl} = \begin{cases} 0, & r_{jl} = e_{il}, \\ r_{jl} - e_{il}, & r_{jl} \neq e_{il}, \end{cases} \quad i \in M, j \in N, l \in K. \tag{4}$$

(3) When tenant T_i has a cost-type requirement of attribute C_l, the lower the value of the cost attribute, the better. d_{ijl} is calculated by the following three cases:

Case 1: when $e_{il} = r_{jl}$, the tenant's T_i performance to landlord L_j is neither a gain nor a loss given attribute C_l.

Case 2: when $e_{il} < r_{jl}$, the tenant's T_i performance to landlord L_j is a loss given attribute C_l. The loss is $e_{il} - r_{jl}$.

Case 3: when $e_{il} > r_{jl}$, the tenant's T_i performance to landlord L_j is a gain given C_l. The gain is $e_{il} - r_{jl}$.

Therefore, when attribute C_l is a cost-type requirement, the equation for calculating d_{ijl} is as follows:

$$d_{ijl} = \begin{cases} 0, & r_{jl} = e_{il}, \\ e_{il} - r_{jl}, & r_{jl} \neq e_{il}, \end{cases} \quad i \in M, j \in N, l \in K. \tag{5}$$

Similarly, landlord L_j should consider whether the tenant's T_i expected rent is within an acceptable range and meets the requirements for the ORH's area. If $F_l = [a_{ijl}]_{m\times n}$ is the profit and loss matrix of landlord L_j for attribute C_l and a_{ijl} is the profit and loss value of landlord L_j with respect to tenant T_i for attribute C_l, then the calculation equation of a_{ijl} is as follows:

$$a_{ijl} = \begin{cases} 0, & r_{jl} = e_{il}, \\ \dfrac{p_{ij}}{r_{jl}} - 1, & e_{il} < p_{ij} < r_{jl}, \\ 1, & r_{jl} < p_{ij} < e_{il}, \end{cases} \quad i \in M, j \in N, l \in K, \tag{6}$$

$$a_{ijl} = \begin{cases} 1, & e_{il} = r_{il}, \\ 0, & e_{il} \neq r_{jl}, \end{cases} \quad i \in M,\ j \in N,\ l \in K. \tag{7}$$

The SAD sides concern about some attributes of online rental house. As the dimensions of these attributes are different, $D_l = [d_{ijl}]_{m\times n}$ and $F_l = [a_{ijl}]_{m\times n}$ of the profit and

loss matrices of both parties in the transaction were converted into normative $D_l' = [d_{ijl}']_{m\times n}$ and $F_l' = [a_{ijl}']_{m\times n}$, among which the calculation equations of d_{ijl}' and a_{ijl}' are as follows [43]:

$$d_{ijl}' = \frac{d_{ijl}}{\max\left|d_{ijl}\right|}, \quad i \in M,\ j \in N,\ l \in K, \tag{8}$$

$$a_{ijl}' = \frac{a_{ijl}}{\max\left|a_{ijl}\right|}, \quad i \in M,\ j \in N,\ l \in K. \tag{9}$$

3.2.2. Basis of the Construction of Perceptual Utility Matrices. Based on the disappointment theory, the normalized profit and loss matrices $D_l' = [d_{ijl}']_{m\times n}$ and $F_l' = [a_{ijl}']_{m\times n}$ are transformed into the perceived utility matrices $V_l^b = [v_{ijl}^b]_{m\times n}$ and $V_l^s = [v_{ijl}^s]_{m\times n}$, and the perceived utility value of both supplier and the demander is determined by the disappointment and delight function $\varphi(x)$ of Grant and Kajii [44], which can be expressed as [26]

$$\varphi(x) = \begin{cases} 1-\alpha^x, & x \geq 0, \\ \beta^{(-x)} - 1, & x < 0. \end{cases} \tag{10}$$

α is the delight parameter satisfying $0 < \alpha < 1$, while β is the disappointment avoidance parameter satisfying $0 < \beta < 1$. The larger α and β are, the lower the subject's perceived value of the matching results is compared to the expected value. In order to facilitate calculation, Laciana and Weber measured the beta value in accordance with most subject behavioral preferences [45], $0.7 \leq \alpha \leq 0.9, 0.7 \leq \beta \leq 0.9$, In this paper, α and β take the same value. Therefore, if $\alpha = \beta = 0.8$, then, $V_l^b = [v_{ijl}^b]_{m\times n}$ and $V_l^s = [v_{ijl}^s]_{m\times n}$, and the calculation equations are as follows [26]:

$$v_{ijl}^b = \varphi\left(d_{ijl}'\right), \quad i \in M,\ j \in N,\ l \in K, \tag{11}$$

$$v_{ijl}^s = \varphi\left(a_{ijl}'\right), \quad i \in M,\ j \in N,\ l \in K. \tag{12}$$

The tenant's T_i comprehensive perceptual utility matrix $V = [v_{ij}^b]_{m\times n}$ is constructed according to the tenant's T_i perceptual utility matrix $V_l^b = [v_{ijl}^b]_{m\times n}$, where the calculation equation of v_{ij}^b is as follows [46]:

$$v_{ij}^b = \sum_{l=1}^{k} w_{il} v_{ijl}^b, \quad i \in M,\ j \in N,\ l \in K. \tag{13}$$

Similarly, a comprehensive perceived utility matrix $V\prime = [v_{ij}^s]_{m\times n}$ of landlord L_j is constructed, where the calculation equation of v_{ij}^s is as follows [46]:

$$v_{ij}^s = \sum_{l=1}^{k} w_{jl} v_{ijl}^s, \quad i \in M,\ j \in N,\ l \in K. \tag{14}$$

In the constructed comprehensive perceived utility matrix, the larger v_{ij}^b and v_{ij}^s, the higher the satisfaction of both suppliers and demanders.

3.2.3. Construction of Supply-Demand Matching Model Based on the Perceived Utility. Set x_{ij} represents a 0-1 variable, where $x_{ij} = 0$ represents that tenant T_i and landlord L_j do not match, while $x_{ij} = 1$ represents that tenant T_i and landlord L_j match. According to the comprehensive perceived utility matrices V and V', under the requirement of stable matching, the following two-objective optimization model can be established to maximize the comprehensive perceived utility of suppliers and demanders:

$$\max \quad Z_1 = \sum_{i=1}^{m}\sum_{j=1}^{n} v_{ij}^b x_{ij}, \quad i \in M,\ j \in N, \tag{15a}$$

$$\max \quad Z_2 = \sum_{i=1}^{m}\sum_{j=1}^{n} v_{ij}^s x_{ij}, \quad i \in M,\ j \in N, \tag{15b}$$

$$\text{s.t.} \quad \sum_{i=1}^{m} x_{ij} \leq 1, \quad j \in N, \tag{15c}$$

$$\sum_{j=1}^{n} x_{ij} = 1, \quad i \in M, \tag{15d}$$

$$\sum_{j=1}^{n} x_{ij} r_{ij} = 1, \quad i \in M, \tag{15e}$$

$$x_{ij} = 0 \text{ or } 1, \quad i \in M,\ j \in N. \tag{15f}$$

Equations (15a) and (15b) are the objective functions, respectively, maximizing the sums of the perceived utilities of suppliers and demanders in the final matching result. Equations (15c) and (15d) are the requirements of two-sided matching. Equation (15c) is an inequality requirement because $m \leq n$, which means that each tenant T_i can match at most one landlord L_j. Equation (15d) is an equality constraint, meaning that each landlord L_j can only match one tenant T_i, while equation (15e) is a hard requirement.

In order to realize the multiobjective optimization model and maximize the overall matching degree between suppliers and demanders, the linear weighting method can be used to weigh equations (15a) and (15b). w_1 and w_2 are the weights of Z_1 and Z_2, respectively, to satisfy $0 \leq w_1, w_2 \leq 1$, and $w_1 + w_2 = 1$. Considering the fairness of the suppliers and demanders, this paper provides that $w_1 = w_2 = 0.5$, and the two-objective model (15a)–(15f) can be transformed into a single-objective optimization model (16a)–(16e):

$$\max \quad Z = 0.5Z_1 + 0.5Z_2, \tag{16a}$$

$$\text{s.t.} \quad \sum_{i=1}^{m} x_{ij} \leq 1, \quad j \in N, \tag{16b}$$

$$\sum_{j=1}^{n} x_{ij} = 1, \quad i \in M, \tag{16c}$$

$$\sum_{j=1}^{n} x_{ij} r_{ij} = 1, \quad i \in M, \tag{16d}$$

$$x_{ij} = 0 \text{ or } 1, \quad i \in N,\ j \in M. \tag{16e}$$

This single-objective optimization model will attempt to apply the improved genetic algorithm given the large scale of this study's two-sided matching and requirements [47, 48].

3.2.4. Solution Procedure. According to the above analysis, the proposed method's solution procedure for the multiattribute ORH SAD matching decision model is summarized as follows:

Step 1: collection, selection, and measurement of Shanghai's housing property SAD information from its online rental platforms using a web crawler

Step 2: calculation of the profit and loss matrices of SAD based on equations (1)–(7)

Step 3: calculation of the normative profit and loss matrices of SAD based on equations (8) and (9)

Step 4: calculation of the perceived utility matrix of SAD based on equations (11) and (12)

Step 5: calculation of the comprehensive perceived utility matrix of SAD based on equations (13) and (14) followed by the conversion of the multiobjective optimization model to a single-objective optimization model based on the comprehensive perceived utility matrix in equations (15a)–(15f) and (16a)–(16e)

Step 6: solving the optimization model in equations (16a)–(16e) to obtain the optimal matching pairs of online rental platforms' SAD sides using a genetic algorithm

3.3. Data and Measurements. This section includes three sections. First, we introduce the background of the Shanghai rental housing market and explain why it is suitable for testing our study. Second, we introduce the data set, which is from Shanghai's current online rental platforms. Third, we introduce the selection and measurement of multiattribute indicators.

3.3.1. Background of Shanghai's Rental Housing Market. Shanghai is one of China's four municipalities. It is located in the east of China at the mouth of the Yangtze River and is the leading city in the Yangtze River Economic Belt. In 2017, Shanghai was organized by 16 districts and counties as shown in Figure 3, including Huangpu, Xuhui, Changning, Jingan, Hongkou, Yangpu, Minhang, Baoshang, Jiading, and Pudong. By the end of 2018, the total population of Shanghai had reached 24.24 million, of which 9.76 million was the floating population, ranking the highest in the country.

Since 1980, China's housing market has been dominated by sales, while the rental market developed slowly. However, with the acceleration of China's urbanization process—a powerful means of regulating urban real estate, retaining the working population, and attracting new talents—the rental housing market has become an important part of deepening the reform of the housing system and an important way to achieve the goal of improving urban residents' quality of life. Shanghai, as China's largest economic center city, attracts a large floating population. Since 2011, Shanghai's permanent migrant population has remained above 9.6 million per year, which accounts for more than 40% of the total resident population, as shown in Figure 4. This huge population size has resulted in an increase in living demands. Therefore, the development of Shanghai's rental housing market requires urgent attention.

In July 2017, China's Ministry of Housing and Construction and nine other departments jointly issued the "Notice on Accelerating the Development of Housing Rental Market in Large- and Medium-sized Cities with Net Population Inflow," requiring large- and medium-sized cities with a net inflow of population to set up government-backed rental housing service platforms. Therefore, Shanghai now has a public house leasing service platform. In addition to this, Shanghai has many private online leasing platforms, including Lianjia.com, Anjuke, and Fangtianxia. Many people use these platforms to find listings. Online platforms can display a variety of information about the listed property, such as the number of rooms, the size of the property, the degree of decoration, and the housing facilities. The platforms also include some rent-seeking information, which helps match the SAD sides. Therefore, Shanghai is a suitable sample for testing the model constructed in this study.

3.3.2. Data Collection. Using the Python web crawler technology, on May 13th, 2019, we performed a crawling exercise on 21,762 articles from the current mainstream online rental platforms like Shanghai's housing leasing public service platform, Anjuke, Lianjia, and Fang.com. At the same time, we collected 1,739 and 929 rent-seeking messages from 58.com and Weibo, respectively, and obtained 1,226 effective massages after manually deleting unrelated data. The keywords used for detecting unrelated data included "I am an owner," "I am an intermediary," "commercial," "shop," "rent out," "office," and "office buildings." The origins of the final data are spread across 15 of Shanghai's 16 districts. The workflow used to crawl rent-seeking information on 58.com is shown in Figure 5. According to the crawled rent-seeking information shown in Table 2, there is a high degree of similarity between the housing attribute indicators considered by the potential tenants. The frequencies of the feature words related to housing area, decoration, rental price, appliances, transportation, and so on, are relatively high. Given that Shanghai's floating population has a relatively high demand for rental housing and that the consideration factors are relatively complete in the transaction, it appears that the collected rent-seeking information is adequately representative, with a strong reference for other cities as well.

3.3.3. Selection and Measurements of Multiattribute Indicators. Word frequency can reflect the importance of a word in different contexts or expressions. The higher the

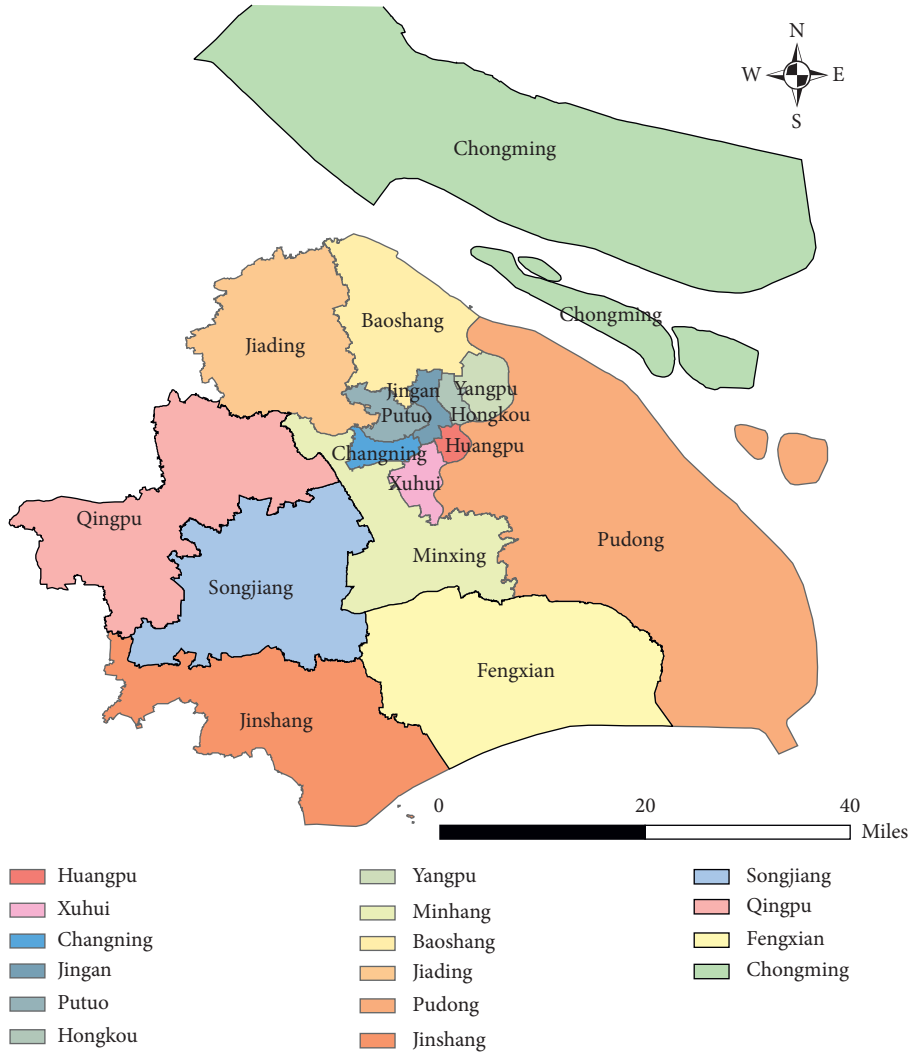


Figure 3: Administrative divisions of Shanghai in 2017.

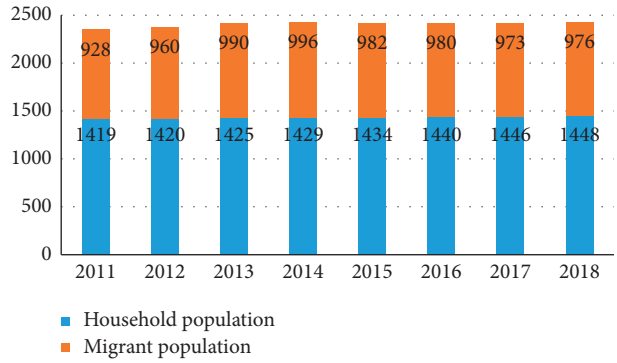


Figure 4: Population structure of Shanghai's permanent residents, 2011–2018 (unit: 10,000).

word frequency, the more important it is. With the aforementioned Python crawled rent-seeking messages from the "Personalized needs" page of 58.com and the comments from Weibo's "Shanghai Rent-seeking" page, this paper employs a word cloud to display the top 50 words related to property characteristics (shown in Figure 6). The top three words are "near," "rent," and "decoration" followed by "subway," "furniture," "home appliances," and "single room." The search term "Near" reflects the requirements of the housing location, while "Subway" indicates a requirement for proximity to public transportation. The characteristic value of "rent" reflects the demand for renting a house. "Room"

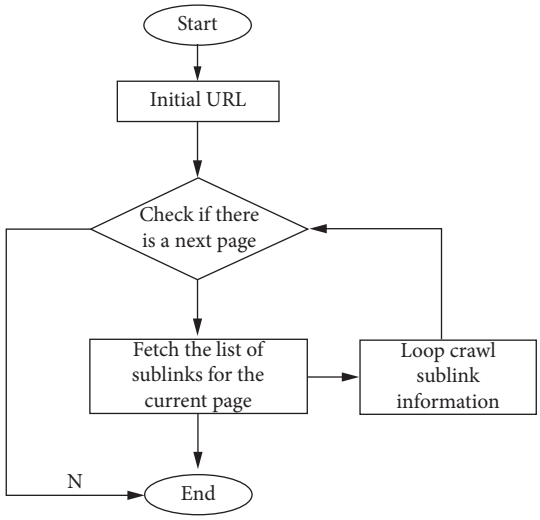


Figure 5: Workflow for crawling rent-seeking information on 58.com.

Table 2: Tenant's rent-seeking information: frequency of feature words (top 30).

Feature words	Frequency
Near	186
House	137
Decoration	124
Clean	113
Furniture	109
Subway	106
Rent	94
One room	94
Home appliances	93
Traffic	80
Hardcover	80
Personal	78
Subway station	67
Price	63
Cooking	63
Tidy	60
Within	60
Share	57
Two rooms	57
Paperback	56
Landlord	56
Elevator	53
Consider	52
Budget	52
Distance	52
Facility	52
Apartment	48
Traffic	45
Complete	41
Health	41

indicates a requirement for a living room; "Decoration," "Furniture," and "appliances" are requirements for the aesthetics, furniture, and amenities of the house.

Although a word frequency analysis can reflect the personalized needs of the tenant by extracting high-frequency words, it cannot reflect the connections between those words. However, a semantic network analysis can visually display the relationship between elements by constructing a semantic relationship network map [49]. Therefore, for this study, we performed a text analysis of the crawled information using ROSTCM6 software. First, the crawled rent-seeking information was processed through word segmentation. Second, the high-frequency words were extracted, and the meaningless words were filtered, then the line feature words were extracted to generate VNA files, and we then imported them into NetDraw software to construct a co-occurrence matrix. Finally, the semantic relationship network map (shown in Figure 7) of the rent-seeking information was generated. As seen in Figure 7, "near," "rent," "decoration," "appliances," and "subway" are the central nodes of the entire network, reflecting the similarity of different tenants in selecting online rental houses. Comprehensively, tenants conduct a systematic investigation of overall quality when searching rental houses. On the one hand, in terms of location and proximity to transportation, tenants share similar requirements, reflected by the "subway," "traffic," and "near" nodes. On the other hand, in terms of the overall quality of the houses, tenants require properties with facilities, as reflected by the "appliances," "furniture," and "kitchen" nodes. In fact, these characteristic words reflect the real needs of tenants when choosing a rental house online. However, both SAD sides of ORH include institutions and/or individuals; thus, the transaction takes place between unfamiliar entities. In such a case, the authenticity of housing and the construction of the social credit system are significant for the smooth operation of platforms [50]. Therefore, this study analyzed the word cloud maps and semantic network maps and finally selected seven attributes: district, traffic convenience, rent, living room, house supporting facilities, decoration, and online ratings. These seven indicators of the crawled listing and rent-seeking information were then quantified separately, as shown in Table 3.

The entropy weight method is effective for calculating the weight and objective orderings of multiattribute indicators by applying information entropy [51]. The smaller the information entropy of the indicator, the larger the amount of information provided. Similarly, the greater the entropy's role in the comprehensive evaluation, the higher the weight. Thus, the paper uses the entropy weight method to calculate weight. For the tenants, it is necessary to pay attention to the seven attributes of area, transportation convenience, rent, living room, housing matching, decoration, and online ratings. Regarding their types, location is a hard requirement, living room, house matching, and decoration are benefit requirements, while rent and transportation convenience are cost requirements. Their weight values are $w_1 = 0.302$, $w_2 = 0.137$, $w_3 = 0.193$, $w_4 = 0.074$, $w_5 = 0.162$, $w_6 = 0.038$, and $w_7 = 0.094$. For the landlords, the main focus is location and rent. In this paper, the weight values of both attributes are 0.5 for fairness.

4. Results and Discussion

4.1. Results. For the construction of the model, first, we randomly selected 50 of the 1,226 tenants whose information

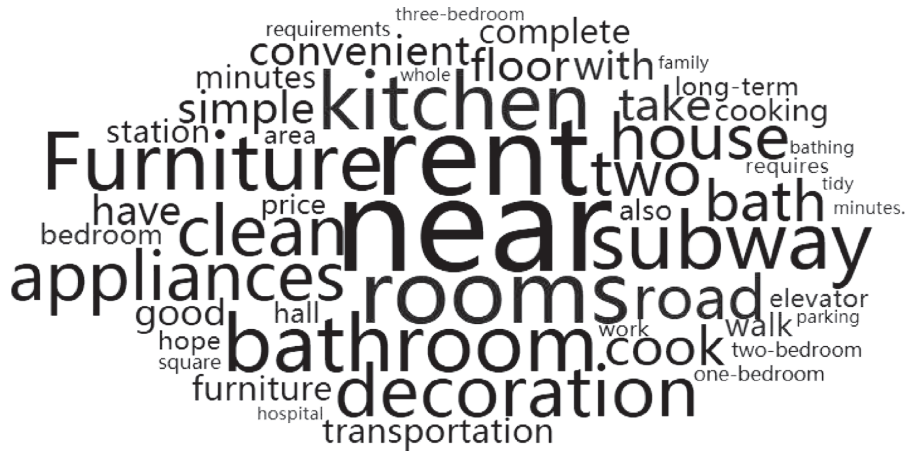


Figure 6: Word cloud diagram of rent-seeking information results.

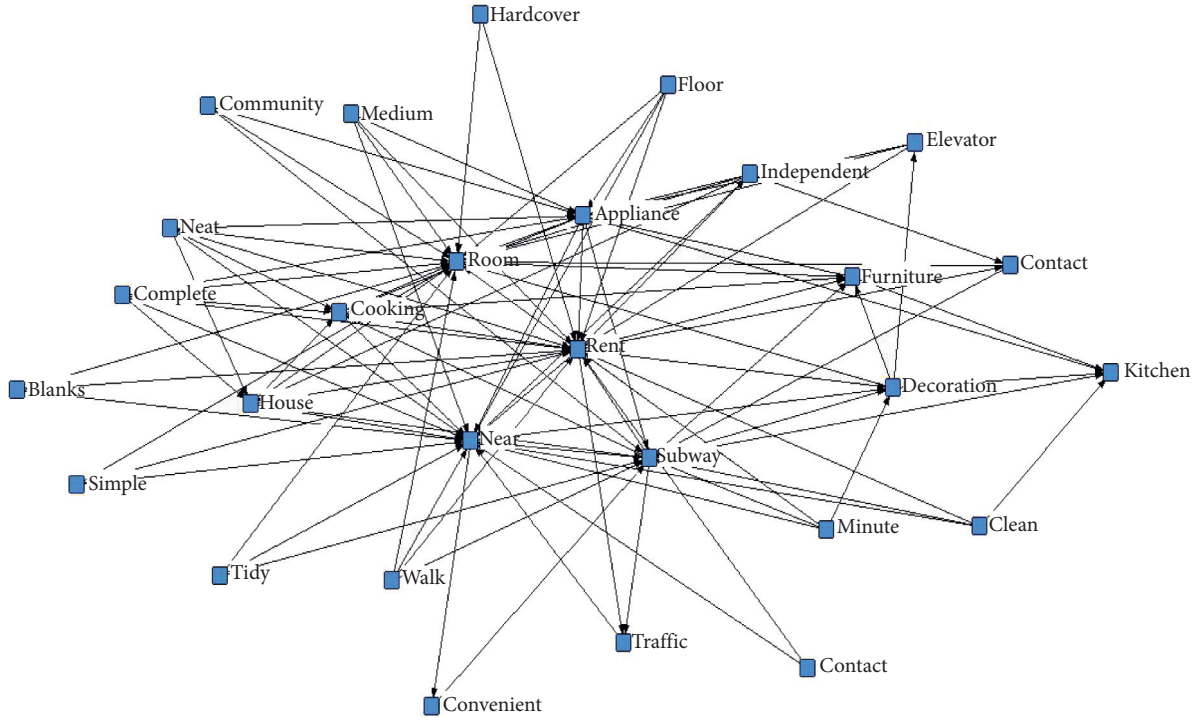


Figure 7: Semantic relationship network map of rent-seeking information.

was crawled; then, according to their needs, 50 landlords were selected. Second, each attribute index was measured. The landlord's evaluation values of the housing attributes are shown in Table 4, while the tenants' expectation values of housing attributes are shown in Table 5. Third, the profit and loss matrices, the standardized profit and loss matrices, the perceived utility matrices, and the comprehensive perceived utility matrices of the SAD were calculated using Matlab 2016 software. The specific steps of the calculations are as follows:

Step 1: according to Table 3, the landlords' evaluation values of the housing attributes and the tenants' expectation values of the housing attributes were measured (as shown in Tables 4 and 5, respectively).

Step 2: by using equations (1)–(7), the landlords' profit and loss matrices $F_l = [a_{ijl}]_{50\times50}$ and the tenants' profit and loss matrices $D_l = [d_{ijl}]_{50\times50}$ for location and transportation convenience were calculated (as shown in Tables 6 and 7 and Tables 8 and 9, respectively). The remaining tenants' profit and loss matrices are shown in Appendix.

Step 3: based on the landlords' and tenants' profit and loss matrices and by using equations (8) and (9), the landlords' normative profit and loss matrices $F_l' = [a_{ijl}']_{50\times50}$ were calculated (as shown in Tables 10 and 11), and the tenants' profit and loss matrices $D_l' = [d_{ijl}']_{50\times50}$ for location and transportation convenience were also calculated (as shown in Tables 12 and 13). The

Table 3: Measurement of multiattribute indicators.

Variable name	Original value of the variable	Variable type	Quantitative value of the variable
Location (C_1)	Huangpu, Jingan, Xuhui, Pudong, Changning, Hongkou, Yangpu, Putuo, Minhang, Baoshang, Jiading, Qingpu, Fengxian, Jinshan, Songjiang	Character classification	Assign 1 to 15, respectively
Transportation convenience (C_2)	It is (1) within 500 meters from the subway, (2) between 500 and 1000 meters away, (3) 1000–1500 meters away, (4) 500–2000 meters away, and (5) 2000 meters and above away.	Character classification	Assign 1 to 5, respectively
Rent (C_3)	Rent value	Numerical quantification	—
Living room (C_4)	One room to four or more rooms	Character classification	Assign 1 to 5, respectively
Appliances (C_5)	No, simple furniture, home appliances, basic configuration, complete equipment, full-fit	Character classification	Assign 1 to 5, respectively
Decoration degree (C_6)	Blank, simple decoration, medium decoration, fine decoration, luxury decoration	Character classification	Assign 1 to 5, respectively
Online ratings (C_7)	Low, average, good, excellent, excellent	Character classification	Assign 1 to 5, respectively

Table 4: Landlords' evaluation values of housing attributes.

L_n	C_1	C_2	C_3	C_4	C_5	C_6	C_7
L_1	1	2	4300	1	4	4	4
L_2	1	2	4000	1	4	3	5
L_3	2	2	3200	1	4	4	4
L_4	2	2	3000	1	3	3	3
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
L_{47}	10	1	6500	3	4	4	4
L_{48}	11	2	5800	3	4	4	3
L_{49}	14	4	3200	3	4	4	3
L_{50}	15	4	6500	3	4	4	4

Table 5: Tenants' expectation values of housing attributes.

T_m	C_1	C_2	C_3	C_4	C_5	C_6	C_7
T_1	1	1	4000	1	4	4	4
T_2	1	2	3800	1	3	3	3
T_3	1	2	6500	2	4	3	2
T_4	2	2	3000	1	3	4	3
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
T_{47}	14	5	1500	2	2	2	2
T_{48}	15	3	2200	1	4	3	2
T_{49}	15	3	3500	2	4	2	3
T_{50}	15	3	6500	3	4	3	5

Table 6: Profit and loss matrix $F_1 = [a_{ij1}]_{50\times50}$ for landlords' housing location.

a_{ij1}	L_1	L_2	L_3	⋯	L_{48}	L_{49}	L_{50}
T_1	1	1	0	⋯	0	0	0
T_2	1	1	0	⋯	0	0	0
T_3	1	1	0	⋯	0	0	0
⋮	⋮	⋮	⋮	⋱	⋮	⋮	⋮
T_{48}	0	0	0	⋯	0	0	1
T_{49}	0	0	0	⋯	0	0	1
T_{50}	0	0	0	⋯	0	0	1

Table 7: Profit and loss matrix $F_2 = [a_{ij2}]_{50\times50}$ for landlords' housing rent.

a_{ij2}	L_1	L_2	L_3	⋯	L_{48}	L_{49}	L_{50}
T_1	−0.0349	0.0000	0.0000	⋯	−0.1552	0.0000	−0.1923
T_2	−0.0581	−0.0250	0.0000	⋯	−0.1724	0.0000	−0.2077
T_3	0.0000	0.0000	0.0000	⋯	0.0000	0.0000	0.0000
⋮	⋮	⋮	⋮	⋱	⋮	⋮	⋮
T_{48}	−0.2442	−0.2250	−0.1563	⋯	−0.3103	−0.1563	−0.3308
T_{49}	−0.0930	−0.0625	0.0000	⋯	−0.1983	0.0000	−0.2308
T_{50}	0.0000	0.0000	0.0000	⋯	0.0000	0.0000	0.0000

Table 8: Profit and loss matrix $D_1 = [d_{ij1}]_{50\times50}$ for tenants' housing location.

d_{ij1}	L_1	L_2	L_3	⋯	L_{48}	L_{49}	L_{50}
T_1	1	1	0	⋯	0	0	0
T_2	1	1	0	⋯	0	0	0
T_3	1	1	0	⋯	0	0	0
⋮	⋮	⋮	⋮	⋱	⋮	⋮	⋮
T_{48}	0	0	0	⋯	0	0	1
T_{49}	0	0	0	⋯	0	0	1
T_{50}	0	0	0	⋯	0	0	1

Table 9: Profit and loss matrix $D_2 = [d_{ij2}]_{50\times50}$ for tenants' transportation convenience.

d_{ij2}	L_1	L_2	L_3	⋯	L_{48}	L_{49}	L_{50}
T_1	−1	−1	−1	⋯	−1	−3	−3
T_2	0	0	0	⋯	0	−2	−2
T_3	0	0	0	⋯	0	−2	−2
⋮	⋮	⋮	⋮	⋱	⋮	⋮	
T_{48}	1	1	1	⋯	1	−1	−1
T_{49}	1	1	1	⋯	1	−1	−1
T_{50}	1	1	1	⋯	1	−1	−1

Table 10: Normative profit and loss matrix $F_1' = [a_{ij1}']_{50\times50}$ for landlords' housing location.

a_{ij1}'	L_1	L_2	L_3	$\cdots$	L_{48}	L_{49}	L_{50}
T_1	1	1	0	$\cdots$	0	0	0
T_2	1	1	0	$\cdots$	0	0	0
T_3	1	1	0	$\cdots$	0	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
T_{48}	0	0	0	$\cdots$	0	0	1
T_{49}	0	0	0	$\cdots$	0	0	1
T_{50}	0	0	0	$\cdots$	0	0	1

Table 11: Normative profit and loss matrix $F_2' = [a_{ij2}']_{50\times50}$ for landlords' housing rent.

a_{ij2}'	L_1	L_2	L_3	$\cdots$	L_{48}	L_{49}	L_{50}
T_1	−0.0349	0.0000	0.0000	$\cdots$	−0.1552	0.0000	−0.1923
T_2	−0.0581	−0.0250	0.0000	$\cdots$	−0.1724	0.0000	−0.2077
T_3	0.0000	0.0000	0.0000	$\cdots$	0.0000	0.0000	0.0000
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
T_{48}	−0.2442	−0.2250	−0.1563	$\cdots$	−0.3103	−0.1563	−0.3308
T_{49}	−0.0930	−0.0625	0.0000	$\cdots$	−0.1983	0.0000	−0.2308
T_{50}	0.0000	0.0000	0.0000	$\cdots$	0.0000	0.0000	0.0000

Table 12: Normative profit and loss matrix $D_1' = [d_{ij1}']_{50\times50}$ for tenants' housing location.

d_{ij1}'	L_1	L_2	L_3	$\cdots$	L_{48}	L_{49}	L_{50}
T_1	1	1	0	$\cdots$	0	0	0
T_2	1	1	0	$\cdots$	0	0	0
T_3	1	1	0	$\cdots$	0	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
T_{48}	0	0	0	$\cdots$	0	0	1
T_{49}	0	0	0	$\cdots$	0	0	1
T_{50}	0	0	0	$\cdots$	0	0	1

Table 13: Normative profit and loss matrix $D_2' = [d_{ij2}']_{50\times50}$ for tenants' transportation convenience.

d_{ij2}'	L_1	L_2	L_3	$\cdots$	L_{48}	L_{49}	L_{50}
T_1	−0.25	−0.25	−0.25	$\cdots$	−0.25	−0.75	−0.75
T_2	0	0	0	$\cdots$	0	−0.5	−0.5
T_3	0	0	0	$\cdots$	0	−0.5	−0.5
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
T_{48}	0.25	0.25	0.25	$\cdots$	0.25	−0.25	−0.25
T_{49}	0.25	0.25	0.25	$\cdots$	0.25	−0.25	−0.25
T_{50}	0.25	0.25	0.25	$\cdots$	0.25	−0.25	−0.25

remaining tenants' normative profit and loss matrices are given in Appendix.

Step 4: the perceived utility matrices $V_l^s = [v_{ijl}^s]_{50\times50}$ and $V_l^b = [v_{ijl}^b]_{50\times50}$ of the landlords and tenants, respectively, were calculated using equations (11) and (12). The landlords' perceived utility matrices $V_l^s = [v_{ijl}^s]_{50\times50}$ are shown in Tables 14 and 15, while the tenants' perceived utility matrices $V_l^s = [v_{ijl}^s]_{50\times50}$ for house location and transportation convenience are shown in Tables 16 and 17. The remaining perceived utility matrices for the tenants are given in Appendix.

Table 14: Perceived utility matrix $V_1^s = [v_{ij1}^s]_{50\times50}$ for landlords' housing location.

v_{ij1}^s	L_1	L_2	L_3	$\cdots$	L_{48}	L_{49}	L_{50}
T_1	0.2	0.2	0	$\cdots$	0	0	0
T_2	0.2	0.2	0	$\cdots$	0	0	0
T_3	0.2	0.2	0	$\cdots$	0	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
T_{48}	0	0	0	$\cdots$	0	0	0.2
T_{49}	0	0	0	$\cdots$	0	0	0.2
T_{50}	0	0	0	$\cdots$	0	0	0.2

Table 15: Perceived utility matrix $V_2^s = [v_{ij2}^s]_{50\times50}$ for landlords' housing rent.

v_{ij2}^s	L_1	L_2	L_3	$\cdots$	L_{48}	L_{49}	L_{50}
T_1	−0.0078	0.0000	0.0000	$\cdots$	−0.0340	0.0000	−0.0420
T_2	−0.0129	−0.0056	0.0000	$\cdots$	−0.0377	0.0000	−0.0453
T_3	0.0000	0.0000	0.0000	$\cdots$	0.0000	0.0000	0.0000
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
T_{48}	−0.0530	−0.0490	−0.0343	$\cdots$	−0.0669	−0.0343	−0.0712
T_{49}	−0.0205	−0.0138	0.0000	$\cdots$	−0.0433	0.0000	−0.0502
T_{50}	0.0000	0.0000	0.0000	$\cdots$	0.0000	0.0000	0.0000

Table 16: Perceived utility matrix $V_1^b = [v_{ij1}^b]_{50\times50}$ for tenants' housing location.

v_{ij1}^b	L_1	L_2	L_3	$\cdots$	L_{48}	L_{49}	L_{50}
T_1	0.2	0.2	0	$\cdots$	0	0	0
T_2	0.2	0.2	0	$\cdots$	0	0	0
T_3	0.2	0.2	0	$\cdots$	0	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
T_{48}	0	0	0	$\cdots$	0	0	0.2
T_{49}	0	0	0	$\cdots$	0	0	0.2
T_{50}	0	0	0	$\cdots$	0	0	0.2

Table 17: Perceived utility matrix $V_2^b = [v_{ij2}^b]_{50\times50}$ for tenants' transportation convenience.

v_{ij2}^b	L_1	L_2	L_3	$\cdots$	L_{48}	L_{49}	L_{50}
T_1	−0.0543	−0.0543	−0.0543	$\cdots$	−0.0543	−0.1541	−0.1541
T_2	0.0000	0.0000	0.0000	$\cdots$	0.0000	−0.1056	−0.1056
T_3	0.0000	0.0000	0.0000	$\cdots$	0.0000	−0.1056	−0.1056
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
T_{48}	0.0543	0.0543	0.0543	$\cdots$	0.0543	−0.0543	−0.0543
T_{49}	0.0543	0.0543	0.0543	$\cdots$	0.0543	−0.0543	−0.0543
T_{50}	0.0543	0.0543	0.0543	$\cdots$	0.0543	−0.0543	−0.0543

Step 5: the landlords' and tenants' comprehensive perceived utility matrices, $V' = [v_{ij}^s]_{50\times50}$ and $V = [v_{ij}^b]_{50\times50}$, receptively, were calculated using equations (13) and (14), as shown in Tables 18 and 19.

Table 18: Comprehensive perceived utility matrix $V' = [v_{ij}^s]_{50\times 50}$ of landlords.

v_{ij}^s	L_1	L_2	L_3	$\cdots$	L_{48}	L_{49}	L_{50}
T_1	0.0961	0.1000	0.0000	$\cdots$	−0.0170	0.0000	−0.0210
T_2	0.0936	0.0972	0.0000	$\cdots$	−0.0189	0.0000	−0.0226
T_3	0.1000	0.1000	0.0000	$\cdots$	0.0000	0.0000	0.0000
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
T_{48}	−0.0265	−0.0245	−0.0171	$\cdots$	−0.0335	−0.0171	0.0644
T_{49}	−0.0103	−0.0069	0.0000	$\cdots$	−0.0216	0.0000	0.0749
T_{50}	0.0000	0.0000	0.0000	$\cdots$	0.0000	0.0000	0.1000

Table 19: Comprehensive perceived utility matrix $V = [v_{ij}^b]_{50\times 50}$ of tenants.

v_{ij}^b	L_1	L_2	L_3	$\cdots$	L_{48}	L_{49}	L_{50}
T_1	0.0514	0.0544	−0.0074	$\cdots$	0.0088	0.0045	−0.0017
T_2	0.0724	0.0750	0.0148	$\cdots$	0.0294	0.0260	0.0180
T_3	0.0616	0.0621	0.0012	$\cdots$	0.0291	0.0147	0.0209
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
T_{48}	0.0062	0.0094	0.0162	$\cdots$	0.0197	0.0275	0.0658
T_{49}	0.0024	0.0056	0.0073	$\cdots$	0.0211	0.0199	0.0694
T_{50}	−0.0264	−0.0250	−0.0264	$\cdots$	−0.0003	−0.0151	0.0515

According to the landlords' and tenants' comprehensive perceived utility matrices, the multiobjective optimization model, equation (17), was constructed as follows:

$$\begin{aligned}
&\max \quad Z_1 = \sum_{i=1}^{50}\sum_{j=1}^{50} v_{ij}^b x_{ij}, \quad i = 1,2,3,\ldots,50;\ j = 1,2,3,\ldots,50\\
&\max \quad Z_2 = \sum_{i=1}^{50}\sum_{j=1}^{50} v_{ij}^s x_{ij}, i = 1,2,3,\ldots,50;\ j = 1,2,3,\ldots,50\\
&\text{s.t.} \quad \begin{cases}
\sum_{i=1}^{50} x_{ij} \le 1, \quad j = 1,2,3,\ldots,50\\
\sum_{j=1}^{50} x_{ij} = 1, \quad i = 1,2,3,\ldots,50\\
\sum_{j=1}^{50} x_{ij} r_{ij} = 1, \quad i = 1,2,3,\ldots,50\\
x_{ij} = 0 \text{ or } 1, \quad i = 1,2,3,\ldots,50,\ j = 1,2,3,\ldots,50.
\end{cases}
\end{aligned} \tag{17}$$

Transforming the above multiobjective optimization model to a single-objective optimization model (equation (18)) was done as follows:

$$\begin{aligned}
&\max \quad Z = 0.5Z_1 + 0.5Z_2\\
&\text{s.t.} \quad \begin{cases}
\sum_{i=1}^{50} x_{ij} \le 1, \quad j = 1,2,3,\ldots,50\\
\sum_{j=1}^{50} x_{ij} = 1, \quad i = 1,2,3,\ldots,50\\
\sum_{j=1}^{50} x_{ij} r_{ij} = 1, \quad i = 1,2,3,\ldots,50\\
x_{ij} = 0 \text{ or } 1, \quad i = 1,2,3,\ldots,50,\ j = 1,2,3,\ldots,50.
\end{cases}
\end{aligned} \tag{18}$$

Step 6: the optimization model in equation (18) was solved using a genetic algorithm. The derived matching results are shown in Table 20.

As shown in Table 20, all tenants and landlords formed matching pairs. This not only shows matching results that are satisfactory to both sides—with an overall satisfaction value of the matching results of $Z = 4.2374$—but also shows an improvement in matching efficiency. That is, the method reduces matching costs and time for both SAD. Therefore, the SAD matching method for ORH proposed in this study can provide reference to the multiattribute two-sided matching decision method and a basis for cracking the ORH SAD matching problem, which is characterized by multiagents and multilevel supply. The rationality and effectiveness of this method, given the multiple attributes of housing and both SAD sides' perceived utility, have been proved in matching the SAD of ORH.

In this study, a genetic algorithm is used to solve the final matching result of the model. The population evolution curve of the genetic algorithm is shown in Figure 8. As shown in the figure, when the population evolves to 200 generations, the objective function value reaches the optimal value and overall stability. In addition, the total running time of the algorithm is 87.8202 seconds, which is very short. When the scale of the problem is large, it is very feasible to apply genetic algorithms to solve the problem of matching SAD of online rental housing.

4.2. Discussion

4.2.1. Impact of the Weight of SAD. During the transaction between the SAD sides, the weights may not be fair, which may cause the matching result to be inconsistent with the actual situation. To further verify the rationality and comprehensiveness of the algorithm, we verified the weight of the SAD double from 0 to 1, respectively. As shown in Figure 9, the overall satisfaction of the SAD sides increased as the weight of the landlord increased, but the opposite is true for the tenants. This may be because landlords have fewer housing attributes to consider during the transaction matching process.

4.2.2. Impact of the Number of SAD. For online housing platforms, the number of SAD sides may be unequal, and it may be a situation where suppliers exceed demanders. Therefore, to further verify the effectiveness of the model and algorithm, we attempted to calculate a situation in which the suppliers exceed demanders and added 5 pieces of rental information for calculation. It was found that when $m \le n$, the overall satisfaction of both SAD sides became larger, $Z = 4.3832$, which was 0.1458 higher than the satisfaction value when $m = n$. This may be caused by some tenants matching with more suitable landlords, making the satisfaction value higher. The influence of the SAD quantities on the overall satisfaction value Z is shown in Figure 10. In summary, this model is applicable in matching the supply and demand of ORH.

Table 20: Matching results of the multiattribute SAD matching decision model of ORH.

Number	Results
1	(T_1, L_2)
2	(T_2, L_1)
3	(T_3, L_{22})
4	(T_4, L_4)
5	(T_5, L_3)
6	(T_6, L_{23})
7	(T_7, L_5)
8	(T_8, L_{24})
9	(T_9, L_{25})
10	(T_{10}, L_6)
11	(T_{11}, L_{27})
12	(T_{12}, L_{26})
13	(T_{13}, L_{43})
14	(T_{14}, L_7)
15	(T_{15}, L_8)
16	(T_{16}, L_{28})
17	(T_{17}, L_9)
18	(T_{18}, L_{10})
19	(T_{19}, L_{29})
20	(T_{20}, L_{30})
21	(T_{21}, L_{11})
22	(T_{22}, L_{12})
23	(T_{23}, L_{31})
24	(T_{24}, L_{44})
25	(T_{25}, L_{13})
26	(T_{26}, L_{14})
27	(T_{37}, L_{42})
28	(T_{28}, L_{45})
29	(T_{29}, L_{39})
30	(T_{30}, L_{34})
31	(T_{31}, L_{35})
32	(T_{32}, L_{46})
33	(T_{33}, L_{16})
34	(T_{34}, L_{15})
35	(T_{35}, L_{36})
36	(T_{36}, L_{47})
37	(T_{37}, L_{17})
38	(T_{38}, L_{38})
39	(T_9, L_{25})
40	(T_{40}, L_{48})
41	(T_{41}, L_{18})
42	(T_{42}, L_{19})
43	(T_{43}, L_{39})
44	(T_{44}, L_{20})
45	(T_{45}, L_{40})
46	(T_{46}, L_{49})
47	(T_{47}, L_{41})
48	(T_{48}, L_{21})
49	(T_9, L_{25})
50	(T_{50}, L_{50})

Overall, from the analysis of the results, the weights occupied by the SAD affected the overall satisfaction value, but had no effect on the final matching result. In addition, this model is applicable when the supply is equal to the demand or the supply is greater than the demand. This study improves the two-sided matching theory and extends it to the online leasing market to provide matching efficiency between the SAD sides [1].

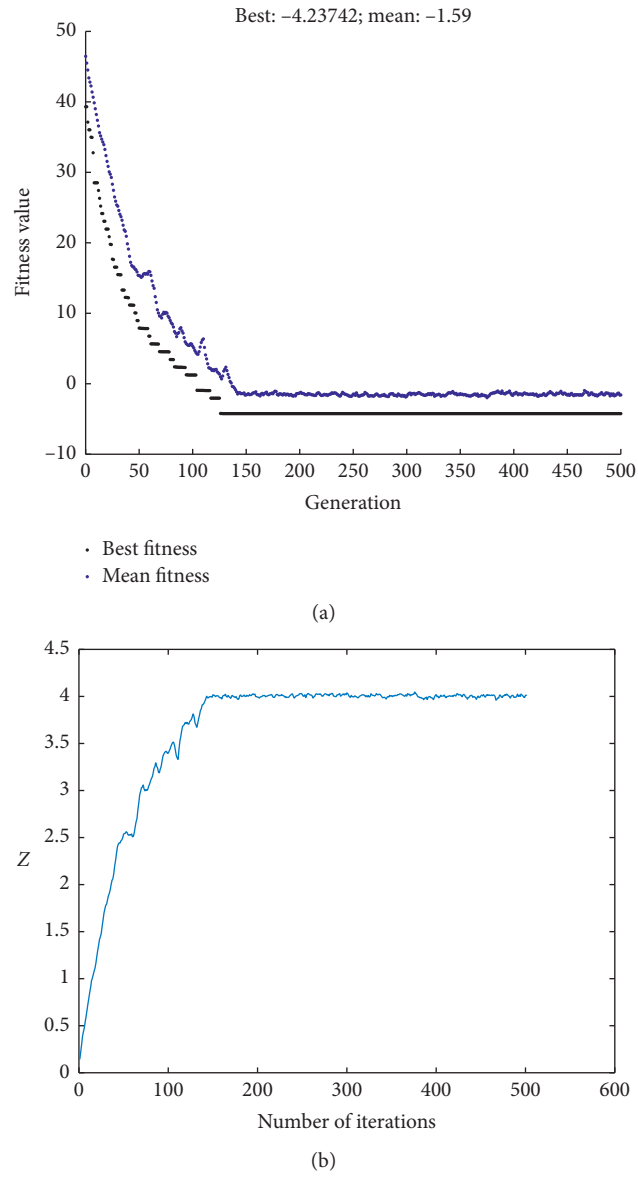


Figure 8: Genetic algorithm evolution curve. (a) Number of iterations. (b) Fitness function.

4.2.3. Comparison with Other Methods. The method proposed in this study may have some differences compared with other methods. With regard to model application, some literature studies have applied matching methods to fields such as hospital diagnosis, carpool matching, job and personnel matching, electronic market matching, electric power retailer and buyer matching, and technology market matching. However, application in the rental housing market is limited. From the perspective of model quantification methods, some scholars quantify the satisfaction values of the SAD based on fuzzy sets, triangular intuitionistic fuzzy numbers, and gray correlation analysis, but these methods do not consider the psychological behavior of participants. Therefore, we quantify the perceived utility value of the SAD sides based on the disappointment theory, starting from the perceptions of the SAD sides, which are

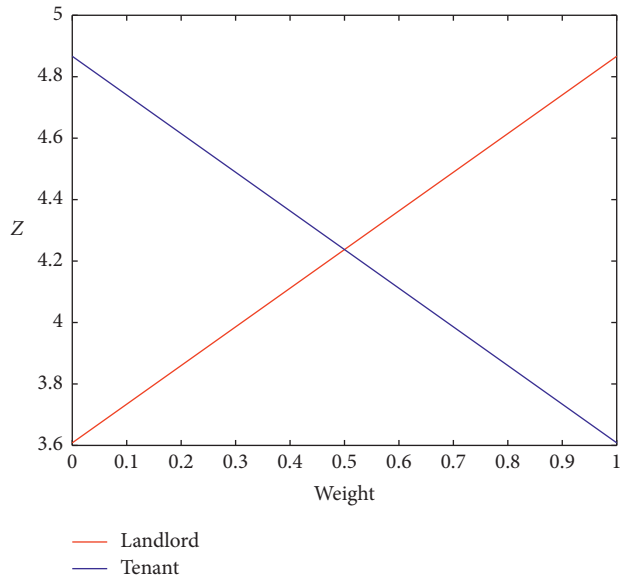


Figure 9: Influence of the weight of the SAD on the overall satisfaction value Z.

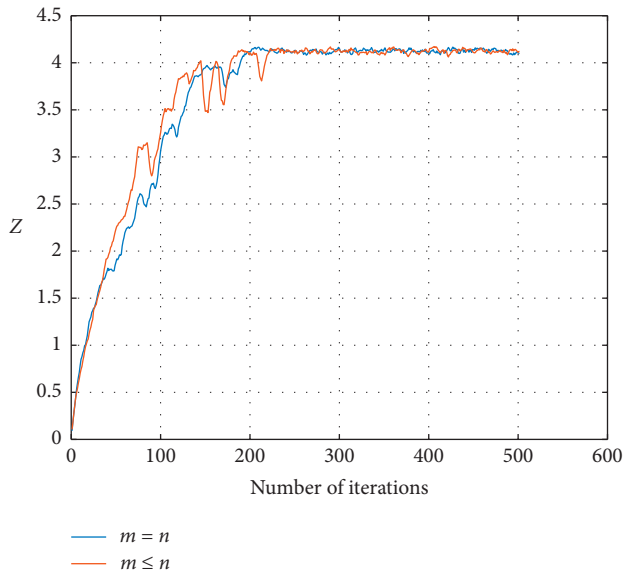


Figure 10: Influence of the number of the SAD on the overall satisfaction value Z.

Table 21: Profit and loss matrix $D_3 = [d_{ij3}]_{50\times50}$ for tenants' housing rent.

d_{ij3}	L_1	L_2	L_3	$\cdots$	L_{48}	L_{49}	L_{50}
T_1	−0.0375	0.0000	0.0000	$\cdots$	−0.2250	0.0000	−0.3125
T_2	−0.0658	−0.0263	0.0000	$\cdots$	−0.2632	0.0000	−0.3553
T_3	0.0000	0.0000	0.0000	$\cdots$	0.0000	0.0000	0.0000
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
T_{48}	−0.4773	−0.4091	−0.2273	$\cdots$	−0.8182	−0.2273	−0.9773
T_{49}	−0.1143	−0.0714	0.0000	$\cdots$	−0.3286	0.0000	−0.4286
T_{50}	0.0000	0.0000	0.0000	$\cdots$	0.0000	0.0000	0.0000

Table 22: Profit and loss matrix $D_4 = [d_{ij4}]_{50\times50}$ for tenants' living room.

d_{ij4}	L_1	L_2	L_3	$\cdots$	L_{48}	L_{49}	L_{50}
T_1	0	0	0	$\cdots$	2	2	2
T_2	0	0	0	$\cdots$	2	2	2
T_3	−1	−1	−1	$\cdots$	1	1	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
T_{48}	0	0	0	$\cdots$	2	2	2
T_{49}	−1	−1	−1	$\cdots$	1	1	1
T_{50}	−2	−2	−2	$\cdots$	0	0	0

Table 23: Profit and loss matrix $D_5 = [d_{ij5}]_{50\times50}$ for tenants' housing matching.

d_{ij5}	L_1	L_2	L_3	$\cdots$	L_{48}	L_{49}	L_{50}
T_1	0	0	0	$\cdots$	0	0	0
T_2	1	1	1	$\cdots$	1	1	1
T_3	0	0	0	$\cdots$	0	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
T_{48}	0	0	0	$\cdots$	0	0	0
T_{49}	0	0	0	$\cdots$	0	0	0
T_{50}	0	0	0	$\cdots$	0	0	0

Table 24: Profit and loss matrix $D_6 = [d_{ij6}]_{50\times50}$ for tenants' housing decoration degree.

d_{ij6}	L_1	L_2	L_3	$\cdots$	L_{48}	L_{49}	L_{50}
T_1	1	0	1	$\cdots$	0	0	0
T_2	1	0	1	$\cdots$	1	1	1
T_3	1	0	1	$\cdots$	1	1	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
T_{48}	1	0	1	$\cdots$	1	1	1
T_{49}	2	1	2	$\cdots$	2	2	2
T_{50}	1	0	1	$\cdots$	1	1	1

Table 25: Profit and loss matrix $D_7 = [d_{ij7}]_{50\times50}$ for tenants' housing online ratings.

d_{ij7}	L_1	L_2	L_3	$\cdots$	L_{48}	L_{49}	L_{50}
T_1	0	1	0	$\cdots$	−1	−1	0
T_2	1	2	1	$\cdots$	0	0	1
T_3	2	3	2	$\cdots$	1	1	2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
T_{48}	2	3	2	$\cdots$	1	1	2
T_{49}	1	2	1	$\cdots$	0	0	1
T_{50}	−1	0	−1	$\cdots$	−2	−2	−1

Table 26: Normative profit and loss matrix $D_3' = [d_{ij3}']_{50\times50}$ for tenants' housing rent.

d'_{ij3}	L_1	L_2	L_3	$\cdots$	L_{48}	L_{49}	L_{50}
T_1	−0.0375	0.0000	0.0000	$\cdots$	−0.2250	0.0000	−0.3125
T_2	−0.0658	−0.0263	0.0000	$\cdots$	−0.2632	0.0000	−0.3553
T_3	0.0000	0.0000	0.0000	$\cdots$	0.0000	0.0000	0.0000
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
T_{48}	−0.4773	−0.4091	−0.2273	$\cdots$	−0.8182	−0.2273	−0.9773
T_{49}	−0.1143	−0.0714	0.0000	$\cdots$	−0.3286	0.0000	−0.4286
T_{50}	0.0000	0.0000	0.0000	$\cdots$	0.0000	0.0000	0.0000

Table 27: Normative profit and loss matrix $D_4' = [d_{ij4}']_{50\times50}$ for tenants' living room.

d_{ij4}'	L_1	L_2	L_3	$\cdots$	L_{48}	L_{49}	L_{50}
T_1	0	0	0	$\cdots$	1	1	1
T_2	0	0	0	$\cdots$	1	1	1
T_3	−0.5	−0.5	−0.5	$\cdots$	0.5	0.5	0.5
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	
T_{48}	0	0	0	$\cdots$	1	1	1
T_{49}	−0.5	−0.5	−0.5	$\cdots$	0.5	0.5	0.5
T_{50}	−1	−1	−1	$\cdots$	0	0	0

Table 28: Normative profit and loss matrix $D_4' = [d_{ij4}']_{50\times50}$ for tenants' housing matching.

d_{ij5}'	L_1	L_2	L_3	$\cdots$	L_{48}	L_{49}	L_{50}
T_1	0.0000	0.0000	0.0000	$\cdots$	0.0000	0.0000	0.0000
T_2	0.3333	0.3333	0.3333	$\cdots$	0.3333	0.3333	0.3333
T_3	0.0000	0.0000	0.0000	$\cdots$	0.0000	0.0000	0.0000
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
T_{48}	0.0000	0.0000	0.0000	$\cdots$	0.0000	0.0000	0.0000
T_{49}	0.0000	0.0000	0.0000	$\cdots$	0.0000	0.0000	0.0000
T_{50}	0.0000	0.0000	0.0000	$\cdots$	0.0000	0.0000	0.0000

Table 29: Normative profit and loss matrix $D_6' = [d_{ij6}']_{50\times50}$ for tenants' housing decoration degree.

d_{ij6}'	L_1	L_2	L_3	$\cdots$	L_{48}	L_{49}	L_{50}
T_1	0.0000	−0.3333	0.0000	$\cdots$	0.0000	0.0000	0.0000
T_2	0.3333	0.0000	0.3333	$\cdots$	0.3333	0.3333	0.3333
T_3	0.3333	0.0000	0.3333	$\cdots$	0.3333	0.3333	0.3333
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
T_{48}	0.3333	0.0000	0.3333	$\cdots$	0.3333	0.3333	0.3333
T_{49}	0.6667	0.3333	0.6667	$\cdots$	0.6667	0.6667	0.6667
T_{50}	0.3333	0.0000	0.3333	$\cdots$	0.3333	0.3333	0.3333

Table 30: Normative profit and loss matrix $D_7' = [d_{ij7}']_{50\times50}$ for tenants' housing online ratings.

d_{ij7}'	L_1	L_2	L_3	$\cdots$	L_{48}	L_{49}	L_{50}
T_1	0.0000	0.3333	0.0000	$\cdots$	−0.3333	−0.3333	0.0000
T_2	0.3333	0.6667	0.3333	$\cdots$	0.0000	0.0000	0.3333
T_3	0.6667	1.0000	0.6667	$\cdots$	0.3333	0.3333	0.6667
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
T_{48}	0.6667	1.0000	0.6667	$\cdots$	0.3333	0.3333	0.6667
T_{49}	0.3333	0.6667	0.3333	$\cdots$	0.0000	0.0000	0.3333
T_{50}	−0.3333	0.0000	−0.3333	$\cdots$	−0.6667	−0.6667	−0.3333

Table 31: Perceived utility matrix $V_3^b = [v_{ij3}^b]_{50\times50}$ for tenants' housing rent.

v_{ij3}^b	L_1	L_2	L_3	$\cdots$	L_{48}	L_{49}	L_{50}
T_1	−0.0083	0.0000	0.0000	$\cdots$	−0.0490	0.0000	−0.0674
T_2	−0.0146	−0.0059	0.0000	$\cdots$	−0.0570	0.0000	−0.0762
T_3	0.0000	0.0000	0.0000	$\cdots$	0.0000	0.0000	0.0000
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
T_{48}	−0.1010	−0.0872	−0.0494	$\cdots$	−0.1669	−0.0494	−0.1959
T_{49}	−0.0252	−0.0158	0.0000	$\cdots$	−0.0707	0.0000	−0.0912
T_{50}	0.0000	0.0000	0.0000	$\cdots$	0.0000	0.0000	0.0000

Table 32: Perceived utility matrix $V_4^b = [v_{ij4}^b]_{50\times50}$ for tenants' living room.

v_{ij4}^b	L_1	L_2	L_3	$\cdots$	L_{48}	L_{49}	L_{50}
T_1	0.0000	0.0000	0.0000	$\cdots$	0.2000	0.2000	0.2000
T_2	0.0000	0.0000	0.0000	$\cdots$	0.2000	0.2000	0.2000
T_3	−0.1056	−0.1056	−0.1056	$\cdots$	0.1056	0.1056	0.1056
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
T_{48}	0.0000	0.0000	0.0000	$\cdots$	0.2000	0.2000	0.2000
T_{49}	−0.1056	−0.1056	−0.1056	$\cdots$	0.1056	0.1056	0.1056
T_{50}	−0.2000	−0.2000	−0.2000	$\cdots$	0.0000	0.0000	0.0000

Table 33: Perceived utility matrix $V_5^b = [v_{ij5}^b]_{50\times50}$ for tenants' housing matching.

v_{ij5}^b	L_1	L_2	L_3	$\cdots$	L_{48}	L_{49}	L_{50}
T_1	0.0000	0.0000	0.0000	$\cdots$	0.0000	0.0000	0.0000
T_2	0.0717	0.0717	0.0717	$\cdots$	0.0717	0.0717	0.0717
T_3	0.0000	0.0000	0.0000	$\cdots$	0.0000	0.0000	0.0000
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
T_{48}	0.0000	0.0000	0.0000	$\cdots$	0.0000	0.0000	0.0000
T_{49}	0.0000	0.0000	0.0000	$\cdots$	0.0000	0.0000	0.0000
T_{50}	0.0000	0.0000	0.0000	$\cdots$	0.0000	0.0000	0.0000

Table 34: Perceived utility matrix $V_6^b = [v_{ij6}^b]_{50\times50}$ for tenants' housing decoration degree.

v_{ij6}^b	L_1	L_2	L_3	$\cdots$	L_{48}	L_{49}	L_{50}
T_1	0.0000	−0.0717	0.0000	$\cdots$	0.0000	0.0000	0.0000
T_2	0.0717	0.0000	0.0717	$\cdots$	0.0717	0.0717	0.0717
T_3	0.0717	0.0000	0.0717	$\cdots$	0.0717	0.0717	0.0717
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
T_{48}	0.0717	0.0000	0.0717	$\cdots$	0.0717	0.0717	0.0717
T_{49}	0.1382	0.0717	0.1382	$\cdots$	0.1382	0.1382	0.1382
T_{50}	0.0717	0.0000	0.0717	$\cdots$	0.0717	0.0717	0.0717

Table 35: Perceived utility matrix $V_7^b = [v_{ij7}^b]_{50\times50}$ for tenants' housing online ratings.

v_{ij7}^b	L_1	L_2	L_3	$\cdots$	L_{48}	L_{49}	L_{50}
T_1	0.0000	0.0717	0.0000	$\cdots$	−0.0717	−0.0717	0.0000
T_2	0.0717	0.1382	0.0717	$\cdots$	0.0000	0.0000	0.0717
T_3	0.1382	0.2000	0.1382	$\cdots$	0.0717	0.0717	0.1382
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
T_{48}	0.1382	0.2000	0.1382	$\cdots$	0.0717	0.0717	0.1382
T_{49}	0.0717	0.1382	0.0717	$\cdots$	0.0000	0.0000	0.0717
T_{50}	−0.0717	0.0000	−0.0717	$\cdots$	−0.1382	−0.1382	−0.0717

highly significant for improving the overall matching satisfaction.

The matching decision method for second-hand housing transactions [41] and the two-sided matching decision method for second-hand housing, which considers the online rental/sale matching attribute association [36], are two relatively close cases. Regarding the matching algorithm, Haaimin [41] designed an extended H-R algorithm that considers that the stability of the matching scheme and the computational workload is large when obtaining matching

results. Wang et al. [36] used the min-max method and uses LINGO software to solve the matching scheme. Although this method is simple and easy to operate, it is not suitable for large-scale two-sided matching problems. For the satisfaction of both SAD, Haaimin [41] ignored the psychological characteristics of misfortunes to avoid losses and maximize advantages. However, the current multiattribute two-sided matching decision model can maximize the satisfaction of SAD. Meanwhile, not only does the genetic algorithm applied in this study solve the large-scale two-sided matching problem but also has a short running time, which can quickly and effectively promote effective matching between SAD.

5. Conclusions and Implication

In this study, an ORH multiattribute SAD matching decision model based on the perceived utility is constructed. This model determines the multiattribute indicators and proportions of housing that both SAD sides are interested in. By considering the psychological expectations of both SAD in measuring the perceived utility values, this model maximizes the overall satisfaction of both SAD. Finally, using Shanghai's online rental and demand information, a genetic algorithm is used to numerically analyze and solve the model, which not only obtains optimal matching results but also proves that the genetic algorithm can solve a large two-sided matching problem. This model enriches the application of double-sided matching theory and provides a reference method for online rental SAD matching. At the same time, this model considers that the psychological perceptions of the SAD sides can maximize their overall satisfaction and promote the matching accuracy of the model.

In the context of "Internet +" and big data, the effective matching of online rental lists is of great implication for promoting the effective operation of the rental housing market. In a theoretical sense, this study enriches the research on rental housing, with particular application for rental platforms, providing a new perspective for the development of rental platforms. In addition, it can be applied to the rental housing markets of other cities. Objectively, the model improves the effective docking of SAD and improves the supply side's efficiency. Subjectively, the enthusiasm for SAD interaction is realized. Therefore, this study has a reference basis for promoting the effective matching of online SAD and the improvement of satisfaction of both SAD.

As with any study, the article at hand suffers from several limitations that are worth considering in further research. Firstly, the method proposed in this study improves the overall matching between the SAD, but it needs to be further compared with other decision-making methods. Such as in the future, we will further study to compare it with other methods that actually produce results. Second, because the economic and social backgrounds of each SAD subject are different, the characteristic variables of the SAD matching decision model in different cities may be different. In the future, we will adjust the multiattribute equations and quantitative standards according to the characteristics of different cities. In addition, the actual practice results should be used to verify the SAD matching decision model in different cities to promote the sustainable development of the rental housing market.

Furthermore, in the transaction process of the ORH platforms, the multiattribute expectations of both parties should be considered. Regardless of the tenants or landlords, each subject considers their own subject's interests based on multiple values, which will directly or indirectly reflect the expected value of the property. Moreover, the supply side of the rental housing market should increase the information symmetry of SAD subjects to ensure that the information is transparent and open and that the transaction process is standardized, guaranteeing easy access to information for both parties.

Appendix

The remaining profit and loss matrices for tenants are given in Tables 21–25, respectively.

The remaining normative profit and loss matrices for tenants are given in Tables 26–30, respectively.

The remaining perceived utility matrices for tenants are given in Tables 31–35, respectively.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This research was supported by the National Key R&D Program of China (grant no. 2018YFD1100202), the Ministry of Education Humanistic and Social Science Program of China (grant no. 19YJC630080), the General Project of Shaanxi Province Soft Science Research Program (grant no. 2019KRM197), the Philosophy and Social Science Research Program of Education Department of Shaanxi Province (grant no.12JK0070), and the Shaanxi Social Science Fund General Project (grant no. 2015R006).

References

[1] A. G. Sharam, "Disruption and the matching market for new multifamily housing in Melbourne, Australia," *Journal of General Management*, vol. 44, no. 3, pp. 160–169, 2019.

[2] A. F. A. Saleh, T. K. Hwa, and R. Majid, "Housing mismatch model in suburban areas," *Procedia—Social and Behavioral Sciences*, vol. 234, pp. 442–451, 2016.

[3] G. Boeing, J. Wegmann, and J. Jiao, "Rental housing spot markets: how online information exchanges can supplement transacted-rents data," *Journal of Planning Education and Research*, 2020.

[4] K. A. Kholodilin, "Quantifying a century of state intervention in rental housing in Germany," *Urban Research & Practice*, vol. 10, no. 3, pp. 267–328, 2017.

[5] O. O. Makinde, "Housing delivery system, need and demand," *Environment, Development and Sustainability*, vol. 16, no. 1, pp. 49–69, 2014.

[6] Z. Z. Jin Lang, "Problems and development countermeasures of China's housing leasing market," *Macroeconomic Management*, vol. 3, pp. 80–85, 2018.

[7] L. Li, L. Feng, X. Guo, H. Xie, and W. Shi, "Complex network analysis of transmission mechanism for sustainable incentive policies," *Sustainability*, vol. 12, no. 2, p. 745, 2020.

[8] C.-H. Chen, F.-J. Hwang, and H.-Y. Kung, "Travel time prediction system based on data clustering for waste collection vehicles," *IEICE Transactions on Information and Systems*, vol. E102.D, no. 7, pp. 1374–1383, 2019.

[9] A. Manduchi, A. Petreski, and A. Stephan, "Market for apartment swap and rental market," *SSRN Electronic Journal*, 2019.

[10] S. Zhou, H. Sun, T. Guan, and T. Li, "Equilibrium model of housing choice for heterogeneous households under public rental housing policy," *Sustainability*, vol. 10, no. 12, p. 4505, 2018.

[11] J. Liu, Y. Zhan, and X. Ma, *Research on Demand of Replacing to Public Rental Housing Tenants in Chongqing*, Springer, Berlin, Germany, 2017.

[12] X. Sun and K. P. Tsang, "What drives the owner-occupied and rental housing markets? Evidence from an estimated DSGE model," *Journal of Money, Credit and Banking*, vol. 49, no. 2-3, pp. 443–468, 2017.

[13] H. Fu, M. Wang, P. Li et al., "Tracing knowledge development trajectories of the Internet of things domain: a main path analysis," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 12, pp. 6531–6540, 2019.

[14] E. Lopez and D. Paredes, "Towards housing policies that consider household's preferences: estimating the demand for housing attributes in Chile," *International Journal of Strategic Property Management*, vol. 24, no. 1, pp. 24–36, 2018.

[15] W. C. Wheaton, "Vacancy, search, and prices in a housing market matching model," *Journal of Political Economy*, vol. 98, no. 6, pp. 1270–1292, 1990.

[16] N. Meng and X. Xu, "Research on customer attention of online short-term rental platform," in *Proceedings of the 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, pp. 2455–2459, IEEE, Chongqing, China, October 2018.

[17] A. Tversky and D. Kahneman, "Loss aversion in riskless choice: a reference-dependent model," *The Quarterly Journal of Economics*, vol. 106, no. 4, pp. 1039–1061, 1991.

[18] D. Kahneman, "A perspective on judgment and choice: mapping bounded rationality," *American Psychologist*, vol. 58, no. 9, pp. 697–720, 2003.

[19] G. Gigerenzer, "Moral satisficing: rethinking moral behavior as bounded rationality," *Topics in Cognitive Science*, vol. 2, no. 3, pp. 528–554, 2010.

[20] J. F. de Medeiros, J. L. D. Ribeiro, and M. N. Cortimiglia, "Influence of perceived value on purchasing decisions of green products in Brazil," *Journal of Cleaner Production*, vol. 110, pp. 158–169, 2016.

[21] D. E. Bell, "Disappointment in decision making under uncertainty," *Operations Research*, vol. 33, no. 1, pp. 1–27, 1985.

[22] G. Loomes and R. Sugden, "Disappointment and dynamic consistency in choice under uncertainty," *The Review of Economic Studies*, vol. 53, no. 2, pp. 271–282, 1986.

[23] A. B. Makar, K. E. McMartin, M. Palese, and T. R. Tephly, "Formate assay in body fluids: application in methanol poisoning," *Biochemical Medicine*, vol. 13, no. 2, pp. 117–126, 1975.

[24] P. Delquié and A. Cillo, "Disappointment without prior expectation: a unifying perspective on decision under risk," *Journal of Risk and Uncertainty*, vol. 33, no. 3, pp. 197–215, 2006.

[25] K. C. Cheung, W. F. Chong, and S. C. P. Yam, "The optimal insurance under disappointment theories," *Insurance: Mathematics and Economics*, vol. 64, pp. 77–90, 2015.

[26] R. Zhao, M. Jin, P. Ren, and Q. Zhang, "Stable two-sided satisfied matching for ridesharing system based on preference orders," *The Journal of Supercomputing*, vol. 76, no. 2, pp. 1063–1081, 2020.

[27] J. Quan, X. Wang, and Y. Quan, "Effects of consumers' strategic behavior and psychological satisfaction on the retailer's pricing and inventory decisions," *IEEE Access*, vol. 7, pp. 178779–178787, 2019.

[28] Z. Zhang, X. Kou, I. Palomares, W. Yu, and J. Gao, "Stable two-sided matching decision making with incomplete fuzzy preference relations: a disappointment theory based approach," *Applied Soft Computing*, vol. 84, Article ID 105730, 2019.

[29] Z.-P. Fan, M.-Y. Li, and X. Zhang, "Satisfied two-sided matching: a method considering elation and disappointment of agents," *Soft Computing*, vol. 22, no. 21, pp. 7227–7241, 2018.

[30] A. E. Roth, "Common and conflicting interests in two-sided matching markets," *European Economic Review*, vol. 27, no. 1, pp. 75–96, 1985.

[31] D. Gale and L. S. Shapley, "College admissions and the stability of marriage," *The American Mathematical Monthly*, vol. 69, no. 1, pp. 9–15, 1962.

[32] F. Kojima and P. A. Pathak, "Incentives and stability in large two-sided matching markets," *American Economic Review*, vol. 99, no. 3, pp. 608–627, 2009.

[33] L. Tan, W. Su, S. Gao, J. Miao, Y. Cheng, and P. Cheng, "Path-flow matching: two-sided matching and multiobjective evolutionary algorithm for traffic scheduling in cloud date center network," *Transactions on Emerging Telecommunications Technologies*, 2019.

[34] C.-H. Chen, "A cell probe-based method for vehicle speed estimation," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E103.A, no. 1, pp. 265–267, 2020.

[35] Y. Lin, Y.-M. Wang, and K.-S. Chin, "An enhanced approach for two-sided matching with 2-tuple linguistic multi-attribute preference," *Soft Computing*, vol. 23, no. 17, pp. 7977–7990, 2019.

[36] R. Wang, D. Li, and G. Yu, "Research on bilateral matching decision method considering attribute association in heterogeneous information environment," *Journal of Intelligent & Fuzzy Systems*, vol. 38, no. 4, pp. 4779–4792, 2020.

[37] T. Gao, M. Huang, Q. Wang et al., "A systematic model of stable multilateral automated negotiation in e-market environment," *Engineering Applications of Artificial Intelligence*, vol. 74, pp. 134–145, 2018.

[38] Y. Kong, W. Huang, and B. Li, "The supply and demand mechanism of electric power retailers and cellular networks based on matching theory," *Information*, vol. 9, no. 8, p. 192, 2018.

[39] B. Xia, S. Shakkottai, and V. Subramanian, "Small-scale markets for a bilateral energy sharing economy," *IEEE Transactions on Control of Network Systems*, vol. 6, no. 3, pp. 1026–1037, 2019.

[40] Y. Yang, S. Luo, J. Fan, X. Zhou, C. Fu, and G. Tang, "Study on specialist outpatient matching appointment and the balance matching model," *Journal of Combinatorial Optimization*, vol. 37, no. 1, pp. 20–39, 2019.

[41] H. M. Liang and Y. P. Jiang, "Decision-making method on second-hand house combination matching," *Systems Engineering—Theory & Practice*, vol. 37, no. 2, pp. 358–367, 2015.

[42] J. Wang, B. Li, W. Li, and S. L. Yang, "Two-sided matching model based on cumulative prospect theory for decision making of two-way referral system," *Journal of Interdisciplinary Mathematics*, vol. 21, no. 5, pp. 1097–1102, 2018.

[43] X. Chen, J. Wang, H. Liang, J. Han, and F. Systems, "Hesitant multi-attribute two-sided matching: a perspective based on prospect theory," *Journal of Intelligent & Fuzzy Systems*, vol. 36, no. 6, pp. 6343–6358, 2019.

[44] S. Grant and A. Kajii, "AUSI expected utility: an anticipated utility theory of relative disappointment aversion," *Journal of Economic Behavior & Organization*, vol. 37, no. 3, pp. 277–290, 1998.

[45] C. E. Laciana and E. U. Weber, "Correcting expected utility for comparisons between alternative outcomes: a unified parameterization of regret and disappointment," *Journal of Risk and Uncertainty*, vol. 36, no. 1, pp. 1–17, 2008.

[46] Y. Liu and K. W. Li, "A two-sided matching decision method for supply and demand of technological knowledge," *Journal of Knowledge Management*, vol. 21, no. 3, pp. 592–606, 2017.

[47] D. G. McVitie and L. B. Wilson, "Stable marriage assignment for unequal sets," *Bit*, vol. 10, no. 3, pp. 295–309, 1970.

[48] C.-H. Chen, F. Song, F.-J. Hwang, L. Wu, and I. Applications, "A probability density function generator based on neural networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 541, Article ID 123344, 2020.

[49] M. L. Doerfel, "What constitutes semantic network analysis? A comparison of research and methodologies," *Connections*, vol. 21, no. 2, pp. 16–26, 1998.

[50] L. Lalicic and C. Weismayer, "The role of authenticity in airbnb experiences," in *Information and Communication Technologies in Tourism*, pp. 781–794, Springer, Cham, Switzerland, 2017.

[51] C. Qiyue, "Structural entropy weight method for determining index weights," *Systems Engineering—Theory & Practice*, vol. 30, pp. 1225–1228, 2017.

Hindawi
Discrete Dynamics in Nature and Society
Volume 2020, Article ID 8586058, 8 pages
https://doi.org/10.1155/2020/8586058

Research Article

Optimizing Ontology Alignment through Improved NSGA-II

Yikun Huang,[1] **Xingsi Xue**,[2,3,4,5] **and Chao Jiang**[2]

[1]*Concord University College, Fujian Normal University, Fuzhou 350118, China*
[2]*College of Information Science and Engineering, Fujian University of Technology, Fuzhou, Fujian 350118, China*
[3]*Intelligent Information Processing Research Center, Fujian University of Technology, Fuzhou 350118, China*
[4]*Guangxi Key Laboratory of Automatic Detecting Technology and Instruments, Guilin University of Electronic Technology, Guilin 541004, China*
[5]*Fujian Key Lab for Automotive Electronics and Electric Drive, Fujian University of Technology, Fuzhou 350118, China*

Correspondence should be addressed to Xingsi Xue; jack8375@gmail.com

Received 18 May 2020; Accepted 1 June 2020; Published 19 June 2020

Guest Editor: Chi-Hua Chen

Over the past decades, a large number of complex optimization problems have been widely addressed through multiobjective evolutionary algorithms (MOEAs), and the knee solutions of the Pareto front (PF) are most likely to be fitting for the decision maker (DM) without any user preferences. This work investigates the ontology matching problem, which is a challenge in the semantic web (SW) domain. Due to the complex heterogeneity between two different ontologies, it is arduous to get an excellent alignment that meets all DMs' demands. To this end, a popular MOEA, i.e., nondominated sorting genetic algorithm (NSGA-II), is investigated to address the ontology matching problem, which outputs the knee solutions in the PF to meet diverse DMs' requirements. In this study, for further enhancing the performance of NSGA-II, we propose to incorporate into NSGA-II's evolutionary process the monkey king evolution algorithm (MKE) as the local search algorithm. The improved NSGA-II (iNSGA-II) is able to better converge to the real Pareto optimum region and ameliorate the quality of the solution. The experiment uses the famous benchmark given by the ontology alignment evaluation initiative (OAEI) to assess the performance of iNSGA-II, and the experiment results present that iNSGA-II is able to seek out preferable alignments than OAEI's participators and NSGA-II-based ontology matching technique.

1. Introduction

Over the past decades, a large number of complex optimization problems have been widely addressed through multiobjective evolutionary algorithms (MOEAs) [1–3]. Generally, the objectives are clashing, preventing concurrent optimization for each objective, which is a challenging and realistic issue [2]. The general approach is to obtain a set of solutions, the so-called Pareto front (PF), that do not dominate each other, and the knee solution of PF is most likely to be fitting for the decision maker (DM) without any user predilections [4, 5]. In this scenario, the problem of ontology alignment in the domain of semantic web (SW) is studied. Ontology, as the kernel technique in SW, can represent a formal definition on the domain knowledge. And, matching ontologies can help find their heterogeneous entities, which can speed up the translation discovery and integrate the knowledge [6]. However, due to the complex heterogeneity in different ontologies, it is arduous to get an excellent alignment to reach all DMs' demand and the ontology matching process usually needs a trade-off between two objectives, i.e., the precision and recall of the obtained alignment.

To meet the diverse requirements of various DMs, a famous multiobjective EA (MOEA), i.e., nondominated sorting genetic algorithm (NSGA-II) [7–10], has been proposed to deal with the ontology matching problem which determines the nondominated solutions and outputs the knee solution of PF as the representation one. Meanwhile, a variety of hybrid optimization approaches have been introduced in the recent years to improve the accuracy and speed of convergence to true optimum solutions which

integrate EA with the local search algorithm (LS) [3, 11–13]. This combination allows gaining high diversity of population due to the high optimization ability which can augment the speed of convergence and reduce the probability of prematurity constringency. In this study, to efficiently determine the knee solution of PF, we propose an improved NSGA-II (iNSGA-II) which incorporates into NSGA-II's evolutionary process the monkey king evolution algorithm (MKE) [14] as the LS algorithm. The proposed iNSGA-II is able to better converge to the real Pareto optimum solution and ameliorate the quality of the solution.

The remainder of the paper is organized as follows: Section 2 demonstrates the related work. Section 3 provides definition of the ontology matching problem and detailed depiction of similarity measures and the aggregation strategy; Section 4 shows the iNSGA-II-based ontology matching technique in detail; Section 5 demonstrates the experimental results; and eventually, Section 6 delineates the conclusion.

2. Related Work

A variety of EA-based matchers with the trait of effectively tackling the ontology alignment problem have been introduced in the recent years. The first category of the EA-based matcher is for the purpose of addressing the problem of ontology metamatching aiming at optimizing the parameters in aggregating diverse ontology matchers' alignment. Naya et al. [15] first proposed the approach by using EA where several similarity measures were combined into one to improve the quality of alignments, and its encoding mechanism is still widely utilized. After that, a memetic algorithm is proposed by Acampora et al. that incorporates the local perturbation into EA to enhance the performance of the algorithm [11]. Acampora et al. [7] and Xue et al. [10] applied the NSGA-II to optimize the alignment whose results were better than the genetic algorithm. Biniz et al. proposed a hybrid approach that combines NSGA-II and a neural network, and its results were effective. Xue et al. [3] proposed that the MatchFmeasure can be approximated to an *f*-measure for better instructing the algorithm's optimization track without using the reference alignment.

The second category of the ontology matching technique based on EA devotes to determine the optimum alignment set of entities. Wang et al. [16] proposed the GAOM (genetic algorithm-based ontology matching) that first models a discrete optimization model for optimizing the mapping set. Chu et al. [17] presented a new metric that took into consideration the entity information in the vector space, and then an EA-based matcher was applied to enhance the alignment quality. A memetic algorithm was proposed by Xue et al., which was effective in instance coreference resolution [12]. Several approaches have been proposed which can efficiently determine the alignments using EA involving user coordination [13, 18, 19]. Xue et al. [20] proposed the CETS (compact evolutionary Tabu search algorithm) to match sensor ontologies, and Acampora et al. [21] compared the performance of five local search algorithms in the ontology matching problem and found that Tabu search results presented the best performance. The proposed iNSGA-II belongs to the first category. Because of the characteristics of exploitation and exploration of MKE, our proposal incorporates NSGA-II, with the best trade-off between precision and recall, and MKE, as a local search algorithm, to further ameliorate the results' quality and reach DMs' demand.

3. Preliminaries

3.1. Ontology Matching Problem. An ontology consists of classes, data type properties, and object properties, which are generally called ontology entities [22]. The purpose of ontology matching is determining the entity correspondence set, which is the so-called ontology alignment. Traditionally, an *f*-measure was often utilized to assess the quality of alignments, which is defined as follows:

$$\begin{aligned} \text{recall} &= \frac{|R \cap A|}{|R|}, \\ \text{precision} &= \frac{|R \cap A|}{|A|}, \\ f - \text{measure} &= 2 \times \frac{\text{recall} \times \text{precision}}{\text{recall} + \text{precision}}, \end{aligned} \tag{1}$$

where R is the reference alignment originated from domain experts and A is the alignment derived from the ontology matcher. Metrics of recall and precision present the alignments' completeness and correctness, which usually balance each other through the reconciliation mean of them, i.e., the alleged *f*-measure. But in the real world, obtaining the reference alignments is particularly expensive, especially when handing the large-scale ontologies. In this paper, we utilize three approximate measures, i.e., MatchCover [3], MeanSimilarity [17], and *f*-measure, to roughly approximate recall, precision, and *f*-measure, respectively. The motivation of MatchCover and MeanSimilarity is to map more identical entities and to ensure the higher similarity of aligned entities, respectively. Assuming the golden alignment's cardinality is one to one, three approximate measures are, respectively, defined as follows:

$$\begin{aligned} \text{MatchCover} &= 2 \times \frac{\left|O_1 - \text{match}\right| + \left|O_2 - \text{match}\right|}{\left|O_1\right| + \left|O_2\right|}, \\ \text{MeanSim} &= \frac{\sum_{i=1}^{|A|} \text{simValue}_i}{|A|}, \\ f - \text{measure}' &= 2 \times \frac{\text{MatchCover} \times \text{MeanSim}}{\text{MatchCover} + \text{MeanSim}}, \end{aligned} \tag{2}$$

where $|O_1|$ and $|O_2|$ are, respectively, the cardinalities of O_1 and O_2; $|O_1 - \text{match}|$ and $|O_2 - \text{match}|$ are the number of matched entities in two ontologies, respectively; $|A|$ is the number of correspondence; $simValue_i$ is the ith entity correspondence's similarity value. Eventually, we define the multiobjective optimization model for the ontology matching problem as follows:

$$\begin{cases} \max & F(\widehat{X}) = \left(f_1(\widehat{X}), f_2(\widehat{X})\right) \\ \text{s.t.} & \widehat{X} = \left(x_1, x_2, \ldots, x_n, x_{n+1}\right)^T, \\ & \sum_{i=1}^{n} x_i = 1, x_i \in [0,1], \quad i = 1, 2, \ldots, n+1, \end{cases} \tag{3}$$

where $f_1(\widehat{X})$ and $f_2(\widehat{X})$ are the MatchCover and Mean-Similarity of $\widehat{X}$, respectively; n is the number of similarity measures utilized; x_i represents the corresponding weight, $i = 1, 2, \ldots, n$; x_{n+1} is the threshold for the purpose of filtering the final alignments.

3.2. Similarity Measures. Typically, similarity measures can be classified into terminology-based, semantic-based, and structure-based measures.

3.2.1. Terminology-Based Measures. Terminology-based measures calculate the string distance between entity identifiers, labels, and comments of two ontologies. There are various terminology-based measures, for instance, Levenshtein distance [23] and Jaro–Winkler distance [24]. In this work, the widely used terminology-based measure is employed, i.e., Levenshtein distance. The Levenshtein distance can be defined as the following equation by the given two strings s_1 and s_2:

$$\text{Levenstein}\left(s_1, s_2\right) = \max\left(0, \frac{\min\left(|s_1|, |s_2|\right) - d\left(s_1, s_2\right)}{\min\left(|s_1|, |s_2|\right)}\right), \tag{4}$$

where $|s_1|$ and $|s_2|$ are, respectively, the string lengths of s_1 and s_2 and $d(s_1, s_2)$ is the quantity of operation necessary to transform s_1 to s_2.

3.2.2. Semantic-Based Measures. The similarity value determined by the semantic-based measures takes into account the semantic information. With the consideration of the entities' identifiers, we employ the WordNet [25], i.e., an electronic vocabulary database that combines different words into a group of synonyms, to calculate the distance based on linguistic relationships, such as synonymy and hypernym. The linguistic distance, $\text{Sim}_{\text{Lin}}(w1, w2)$, is represented as follows by the given two words $w1$ and $w2$:

$$\begin{cases} 1, & \text{if two words are synonymous,} \\ 0.5, & \text{if one word is the hypernym of the other,} \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

Another semantic-based similarity method used is the cosine distance which is expressed by means of the magnitude and dot product from two vectors [26]. Machine learning has a wide range of application scenarios [27–30]; in this study, we employ the trained Word2Vec model (https://mccormickml.com/2016/04/12/googles-pretrained-word2vec-model-in-python/) that trains through a Google News dataset with roughly 100 billion words. Given two entity vectors V_1 and V_2, the cosine distance is defined as follows:

$$\text{cosine distance}\left(V_1, V_2\right) = \frac{V_1 \cdot V_2}{|V_1| \times |V_2|}. \tag{6}$$

3.2.3. Structure-Based Measures. In our study, we compute the structure-based measure that says "elements are similar in two different ontologies if they are related to similar elements." Particularly, the structure-based distance is computed by using the famous algorithm, similarity flooding (SF) [31], where an iterative fixpoint computation (see equation (7)) is used to generate correspondence between the elements of two ontologies:

$$\delta^{i+1} = \text{norm}\left(\delta^i + f\left(\delta^i\right)\right), \tag{7}$$

where δ^i is the value of the last iteration changed in each iteration and f is a function that enables a similarity value of an element pair to increment in accordance with the similarity of its adjacent elements. For more information for the SF algorithm, please see [31].

3.2.4. Aggregation Strategy. A straightforward but practical aggregation strategy is utilized in this study, as an averaged weight method to integrate all mentioned similarity measures above. The method can be described as follows:

$$\text{simValue}\left(e_1, e_2\right) = \sum_{i=1}^{n} w_i \times \text{sim}\left(e_1, e_2\right), \tag{8}$$

where e_1 and e_2 are the entities from different ontologies; n is the number of similarity approaches considered; w_i is the weight of the corresponding similarity measure; $\text{sim}(e_1, e_2)$ is the instance of the represented similarity function above.

4. iNSGA-II for Optimizing Ontology Alignment

In this study, we apply the iNSGA-II to tackle the ontology matching problem. As a flexible and robust approach, NSGA-II can quickly find various nondominated solutions. To further improve the probability of true convergence and enhance the quality of the solutions, as the character of exploitation and exploration of MKE, we introduce the MKE as the local search algorithm. We depict in detail the encoding mechanism, genetic operators, and local search algorithm in the next sections.

4.1. Encoding Mechanism. At the beginning of iNSGA-II, a random parent population is engendered through decimal encoding. The chromosome of an individual can be split into two parts, one representing the several weights and the other representing the threshold. Supposing p is the number of weights we need, then the several cuts could be expressed as $c' = \left\{c'_1, c'_2, \ldots, c'_{p-1}\right\}$. The chromosome decoding is carried out by ordering c' from lower to higher, and then we get $c = \left\{c_1, c_2, \ldots, c_{p-1}\right\}$ and computing the weights as follows:

$$w_k = \begin{cases} c_1, & k = 1, \\ c_k - c_{k-1}, & 1 < k < p, \\ 1 - c_{p-1}, & k = p. \end{cases} \quad (9)$$

Subsequently, the chromosome length is $(n-1)\cdot$cutLength + thresholdLength, where n is the number of required weights and cutLength and thresholdLength are the chromosome lengths of the cut and the threshold, respectively. Figure 1 demonstrates the encoding and decoding mechanisms of several weights.

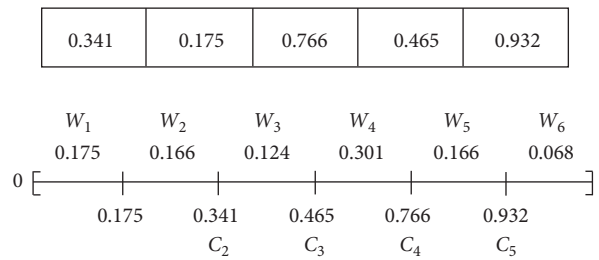


Figure 1: Example showing the encoding and decoding mechanisms of several weights.

4.2. Genetic Operators. The iNSGA-II utilizes the following genetic operators to generate an offspring population.

4.2.1. Selection. The selection operator's target is to select two parents that can be utilized in the crossover operator. In this study, two individuals are selected randomly from parent population as parent_1 and parent_2, respectively.

4.2.2. Crossover. According to the selection operator, we select two chromosomes in the population as parents, i.e., parent_1 and parent_2. We check whether the crossover operator could be applied on the basis of the crossover probability, that is, the parameter in the algorithm. Then, two children can be engendered according to the following formula:

$$\text{child}_i = \text{rand}_i \times \text{parent}_1 + (1 - \text{rand}_i) \times \text{parent}_2, \quad i = 1, 2, \quad (10)$$

where child_i is the i^{th} generated individual and rand_i is a random number in the interval [0, 1].

4.2.3. Mutation. For the purpose of preventing premature convergence and assuring the diversity of population, the mutation operator is utilized. By means of the mutation probability, the newly generated individual can be produced as the following formula:

$$\text{Indi}_{\text{new}} = \begin{cases} \text{rand} \times \text{Indi}_{\text{old}}, & \text{if } r < 0.5, \\ \text{Indi}_{\text{old}} + \text{rand} \times (1 - \text{Indi}_{\text{old}}), & \text{otherwise}, \end{cases} \quad (11)$$

where rand is a random number in the interval [0, 1] and r is a random number to determine whether Indi_{old} should become bigger or not. Furthermore, the formula ensures the newly produced individual in the interval [0, 1].

4.3. Generation of New Population. For the sake of speeding up the algorithm's convergence and ensuring the diversification of population, the generation of a new parent population chooses half of the chromosomes from the front end of the population which combines parent population and offspring population by using nondomination ranks and crowing distance. More details can be seen in [1].

4.4. Local Search Algorithm. After the new parent population is engendered, to efficiently determine the knee solutions of PF, we apply the MKE for local perturbation on each knee solution in the Pareto front when the knee solution does not change in five generations. In this study, the knee solution is the one in the PF with the best recall or precision. In particular, the knee position which is the latent location of the Pareto front presents the greatest trade-off between objectives. For each knee solution, i.e., the best recall and precision, the adopted algorithm for local perturbation in each knee solution is the second version of MKE [14]. The design of MKE is inspired by the Chinese fairy tales, Journey to the West, and the Monkey King which plays a key role in protecting the master and solving various problems in the journey. Equation (12) denotes the equation of $\widehat{X}$ and $\widehat{X}_{\text{knee},G}$, where $\widehat{X}$ is the matrix representing the population, the ith individual X_i is the ith row vector of $\widehat{X}$, s is the scale of population, $X_{\text{knee},G}$ is the knee solution in the population, and $\widehat{X}_{\text{knee},G}$ denotes the knee solution matrix with the $C \times D$ row vector:

$$\widehat{X} = \begin{bmatrix} X_1 \\ X_2 \\ \cdots \\ X_S \end{bmatrix}, \qquad \widehat{X}_{\text{knee},G} = \begin{bmatrix} X_{\text{knee},G} \\ X_{\text{knee},G} \\ \cdots \\ X_{\text{knee},G} \end{bmatrix}_{C \times D}. \quad (12)$$

The formula of updating the knee solution is defined as follows:

$$\begin{cases} \widehat{X}_{\text{diff}} = \left(\widehat{X}_{r1} - \widehat{X}_{r2}\right), \\ \widehat{X}_{\text{mk}} = \widehat{X}_{\text{knee},G} + \text{FC} \times \widehat{X}_{\text{diff}}, \\ X_{\text{knee},G+1} = \text{opt}\left\{X_{\text{mk}}(i), X_{\text{knee},G}\right\}, \end{cases} \quad (13)$$

where $\widehat{X}_{r1} = \begin{bmatrix} X_{r1}^1 \\ X_{r1}^2 \\ \cdots \\ X_{r1}^C \end{bmatrix}_{C \times D}$ and $\widehat{X}_{r2} = \begin{bmatrix} X_{r2}^1 \\ X_{r2}^2 \\ \cdots \\ X_{r2}^C \end{bmatrix}_{C \times D}$, in which C and D are, respectively, the scales of differential individuals

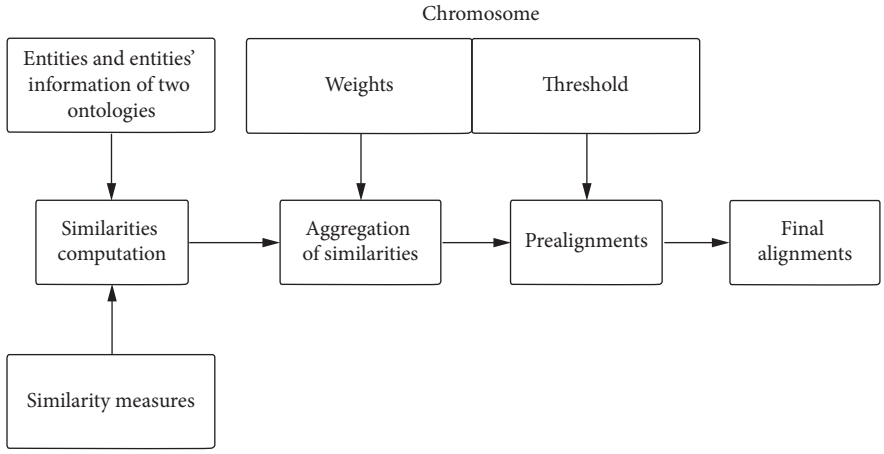


Figure 2: The flow of determining the alignments.

and the cardinality of dimensions; $\widehat{X}_{r1}$ and $\widehat{X}_{r2}$ are randomly produced by choosing the $C \times D$ row vector from the population $\widehat{X}$; $\widehat{X}_{\text{diff}}$ is the difference matrix engendered by the disparity of two random matrices $\widehat{X}_{r1}$ and $\widehat{X}_{r2}$; FC is the fluctuation coefficient of the difference matrices; $X_{\text{mk}}(i)$ is the ith row vector in $\widehat{X}_{\text{mk}}$, $i = 1, 2, \ldots, C$; $X_{\text{knee},G+1}$ is the optimum row vector in $\widehat{X}_{\text{mk}}(i)$ and $X_{\text{knee},G}$, i.e., the individual with the highest fitness. The flow of determining the alignments is shown in Figure 2.

5. Experiment

In this study, the experiment is done by applying the famous benchmark track given by the ontology alignment evaluation initiative (OAEI) (http://oaei.ontologymatching.org/2016/). In the benchmark track, each testing case comprises two to-be-mapped ontologies and the reference alignment, to assess the effectiveness of the ontology matcher. Table 1 provides a concise statement of OAEI's benchmark track.

5.1. Experimental Configuration. In this study, we conduct the experiment by means of the four mentioned similarity measures above, i.e., Levenshtein distance, linguistic distance, cosine distance, and structure distance. The similarity measures should be applied to both input ontologies, and the results are saved in the CSV file before the execution of iNSGA-II. This is done to avoid recalculating the similarities and to boost efficiency of the matching process. The proposed iNSGA-II utilizes the following parameters which provide the trade-off settings acquired empirically to fulfil the highest average quality of alignments across all testing cases in the exploited dataset. Using this parameter configuration, the experiment of this study proves that the selected parameters are robust to all the heterogeneous issues between two ontologies and are expected to be robust for the universal situations of heterogeneity in the real world. Scales of population = 100; iteration for local search = 30; mutation probability = 0.05; crossover probability = 0.7; scales of differential individuals $C = 3$; fluctuation coefficient FC = 0.2; budget condition = 3000 fitness evaluations.

5.2. Experimental Results. For the purpose of comparing the quality of our approach with the NSGA-II-based ontology matcher and OAEI's participators (http://oaei.ontologymatching.org/2016/results/benchmarks/index.html), the obtained alignments have been assessed with conventional recall, precision, and f-measure. The results of NSGA-II and iNSGA-II are mean values of thirty independent runs, and we choose the individual of the best f-measure from knee solutions as a representative. Table 2 displays the mean values and standard deviation of the produced three metrics from our proposal and NSGA-II. Table 3 shows the t-test statistical analysis [32] of values in Table 2. It can be seen from the tables that in all testing cases, as the characteristics of exploitation and exploration of MKE, the quality of alignments obtained by iNSGA-II is better than or equal to NSGA-II.

In Figure 3, X-axis represents different ontology matchers and Y-axis shows the value of f-measure, precision, and recall. Because of iNSGA-II which takes both recall and precision into consideration and the exploitation of the local search, it can be seen from Figure 3 that the f-measure of our proposal is 0.92, which outperforms all the OAEI's participators and NSGA-II in terms of the f-measure. Figure 4 displays that our proposal is more effective than NSGA-II by compared evaluation times of the f-measure value reaching 0.6.

From the above, the quality of alignments of iNSGA-II is better than that of NSGA-II and OAEI's participators, which demonstrates its better trade-off on the algorithm's exploration and exploitation. To conclude, iNSGA-II can efficiently determine high-quality ontology alignments when matching various heterogeneous ontologies.

Table 1: The concise statement for the benchmark track.

ID	Concise statement
101–104	Two identical ontologies
201–210	Two ontologies with different terminology and semantics characteristics
221–247	Two ontologies with different structure characteristics
248–266	Two ontologies with different terminology, semantics, and structure characteristics
301–304	The real-world ontologies

Table 2: Comparison of the *f*-measure, recall, and precision obtained by iNSGA-II and NSGA-II through the mean value and standard deviation.

ID	NSGA-II			iNSGA-II		
	f-measure (st. dev.)	Recall (st. dev.)	Precision (st. dev.)	*f*-measure (st. dev.)	Recall (st. dev.)	Precision (st. dev.)
101	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
103	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
104	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
201	0.94 (0.02)	0.90 (0.01)	0.98 (0.02)	0.95 (0.00)	0.95 (0.01)	0.95 (0.00)
203	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
204	0.98 (0.01)	1.00 (0.00)	0.96 (0.02)	0.99 (0.00)	0.99 (0.00)	0.99 (0.00)
205	0.94 (0.02)	0.91 (0.02)	0.98 (0.01)	0.96 (0.00)	0.96 (0.01)	0.96 (0.01)
206	0.91 (0.01)	0.92 (0.01)	0.91 (0.01)	0.93 (0.00)	0.93 (0.00)	0.93 (0.00)
221	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
222	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
223	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
224	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
225	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
228	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
230	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
231	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
301	0.85 (0.02)	0.81 (0.02)	0.90 (0.01)	0.86 (0.01)	0.82 (0.01)	0.91 (0.01)
302	0.74 (0.01)	0.58 (0.02)	0.97 (0.01)	0.75 (0.00)	0.62 (0.01)	0.95 (0.01)
303	0.87 (0.02)	0.81 (0.01)	0.93 (0.00)	0.87 (0.01)	0.81 (0.00)	0.93 (0.00)
304	0.93 (0.02)	0.89 (0.02)	0.97 (0.01)	0.94 (0.01)	0.91 (0.01)	0.97 (0.01)

Table 3: The analysis of the *t*-test statistical.

ID	*t*-value of *f*-measure (iNSGA-II, NSGA-II)	*t*-value of recall (iNSGA-II, NSGA-II)	*t*-value of precision (iNSGA-II, NSGA-II)
101	0.00	0.00	0.00
103	0.00	0.00	0.00
104	0.00	0.00	0.00
201	2.74	19.36	−8.22
203	0.00	0.00	0.00
204	5.48	0.00	8.22
205	5.48	12.25	−7.75
206	10.95	5.48	10.95
221	0.00	0.00	0.00
222	0.00	0.00	0.00
223	0.00	0.00	0.00
224	0.00	0.00	0.00
225	0.00	0.00	0.00
228	0.00	0.00	0.00
230	0.00	0.00	0.00
231	0.00	0.00	0.00
301	2.45	2.45	3.87
302	5.48	9.80	−7.74
303	0.00	0.00	0.00
304	2.45	4.90	0.00

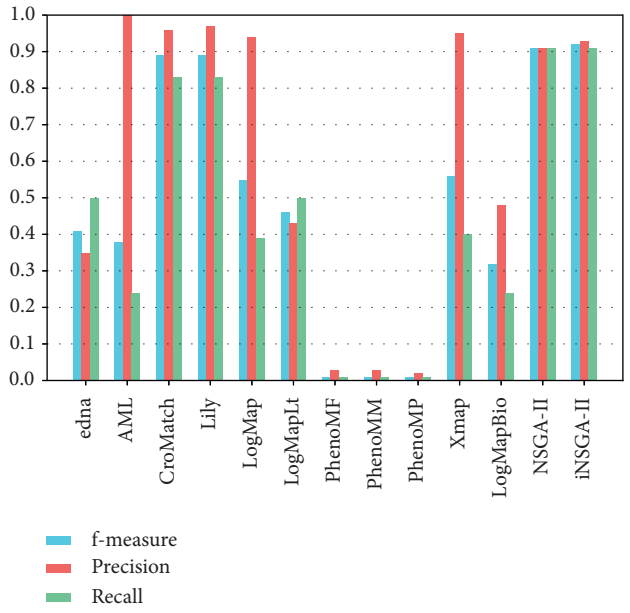


Figure 3: Comparing the results' quality of different ontology matchers.

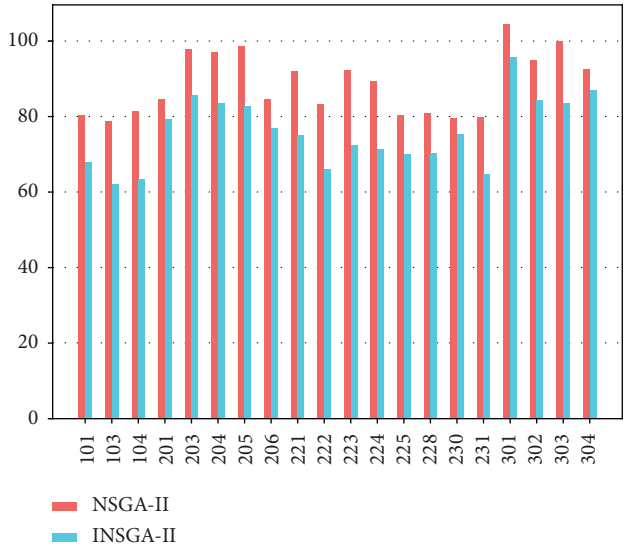


Figure 4: Comparison of evaluation times of the *f*-measure value reaching 0.6.

6. Conclusion

Ontology alignment can find heterogeneous entities of diverse ontologies which can speed up the translation discovery and integrate the knowledge. In this study, the iNSGA-II, which incorporates into NSGA-II's evolutionary process the monkey king evolution algorithm (MKE) as the LS algorithm, is introduced to match the heterogeneous entities which can effectively determine the knee solution of PF enhancing the quality of alignments. By means of the benchmark track of OAEI, the experiment results show that our proposal outperforms the NSGA-II-based ontology matching technique and the OAEI's participators. To further enhance the quality of alignments, in the future, we will focus on semiautomatic ontology matching techniques which introduce interacting process into the MOEA-based matcher, and we will be interested in the word embedding-based similarity measure which would take background knowledge and entities' information of ontologies into consideration to improve the performance of ontology matching [33] techniques.

Data Availability

The data used to support this study can be found in http://oaei.ontologymatching.org.

Conflicts of Interest

The authors declare that they have no conflicts of interest related to this work.

Acknowledgments

This work was supported by the 2019 Fujian Province Undergraduate Universities Teaching Reform Research Project (no. FBJG20190156), the 2018 Program for Outstanding Young Scientific Researcher in Fujian Province University, the Guangxi Key Laboratory of Automatic Detecting Technology and Instruments (no. YQ20206), the Program for New Century Excellent Talents in Fujian Province University (no. GY-Z18155), the Scientific Research Foundation of Fujian University of Technology (no. GY-Z17162), the Science and Technology Planning Project in Fuzhou City (no. 2019-G-40), and the Foreign Cooperation Project in Fujian Province (no. 2019I0019).

References

[1] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.

[2] A. Konak, D. W. Coit, and A. E. Smith, "Multi-objective optimization using genetic algorithms: a tutorial," *Reliability Engineering & System Safety*, vol. 91, no. 9, pp. 992–1007, 2006.

[3] X. Xue and Y. Wang, "Optimizing ontology alignments through a memetic algorithm using both matchfmeasure and unanimous improvement ratio," *Artificial Intelligence*, vol. 223, pp. 65–81, 2015.

[4] S. Bechikh, L. B. Said, and K. Ghedira, "Searching for knee regions of the pareto front using mobile reference points," *Soft Computing*, vol. 15, no. 9, pp. 1807–1823, 2011.

[5] J. Branke, K. Deb, H. Dierolf et al., "Finding knees in multi-objective optimization," in *International Conference on Parallel Problem Solving from Nature*, pp. 722–731, Springer, Berlin, Germany, 2004.

[6] L. Otero-Cerdeira, F. J. Rodríguez-Martínez, and A. Gómez-Rodríguez, "Ontology matching: a literature review," *Expert Systems with Applications*, vol. 42, no. 2, pp. 949–971, 2015.

[7] G. Acampora, U. Kaymak, V. Loia et al., "Applying NSGA-II for solving the ontology alignment problem," in *Proceedings of the 2013 IEEE International Conference on Systems, Man, and*

Cybernetics, pp. 1098–1103, IEEE, Honolulu, HI, USA, July 2013.

[8] M. Biniz and R. El Ayachi, "Optimizing ontology alignments by using Neural NSGA-II," *Journal of Electronic Commerce in Organizations*, vol. 16, no. 1, pp. 29–42, 2018.

[9] X. Xue, J. Chen, J. Chen et al., "A hybrid Nsga-Ii for matching biomedical ontology," in *International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pp. 3–10, Springer, Cham, Switzerland, 2018.

[10] X. Xue, Y. Wang, and W. C. Hao, "Optimizing ontology alignments by using NSGA-II," *International Arab Journal of Information Technology*, vol. 12, no. 2, 2015.

[11] G. Acampora, V. Loia, and A. Vitiello, "Enhancing ontology alignment through a memetic aggregation of similarity measures," *Information Sciences*, vol. 250, pp. 1–20, 2013.

[12] X. Xue and Y. Wang, "Using memetic algorithm for instance coreference resolution," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 2, pp. 580–591, 2016.

[13] X. Xue, J. Chen, and X. Yao, "Efficient user involvement in semiautomatic ontology matching," *IEEE Transactions on Emerging Topics in Computational Intelligence*, pp. 1–11, 2018.

[14] Z. Meng and J.-S. Pan, "Monkey king evolution: a new memetic evolutionary algorithm and its application in vehicle fuel consumption optimization," *Knowledge-Based Systems*, vol. 97, pp. 144–157, 2016.

[15] J. M. V. Naya, M. M. Romero, J. P. Loureiro et al., "Improving ontology alignment through genetic algorithms," *Soft Computing Methods for Practical Environment Solutions: Techniques and Studies*, pp. 240–259, IGI Global, PA, USA, 2010.

[16] J. Wang, Z. Ding, and C. Jiang, "Gaom: genetic algorithm based ontology matching," in *Proceedings of the 2006 IEEE Asia-Pacific Conference on Services Computing (APSCC'06)*, pp. 617–620, IEEE, Guangzhou, China, December 2006.

[17] S. Chu, X. Xue, J.-S. Pan et al., "Optimizing ontology alignment in vector space," *Journal of Internet Technology*, vol. 21, no. 1, pp. 15–23, 2020.

[18] X. Xue and J. Liu, "Collaborative ontology matching based on compact interactive evolutionary algorithm," *Knowledge-based Systems*, vol. 137, pp. 94–103, 2017.

[19] X. Xue and X. Yao, "Interactive ontology matching based on partial reference alignment," *Applied Soft Computing*, vol. 72, pp. 355–370, 2018.

[20] X. Xue and J. Chen, "Using compact evolutionary Tabu search algorithm for matching sensor ontologies," *Swarm and Evolutionary Computation*, vol. 48, pp. 25–30, 2019.

[21] G. Acampora and A. Vitiello, "A study on local search metaheuristics for ontology alignment," in *Computational Intelligence for Semantic Knowledge Management*, pp. 53–70, Springer, Cham, Switzerland, 2020.

[22] J. Euzenat and P. Shvaiko, *Ontology Matching*, Vol. 18, Springer, Berlin, Germany, 2007.

[23] G. Navarro, "A guided tour to approximate string matching," *ACM Computing Surveys (CSUR)*, vol. 33, no. 1, pp. 31–88, 2001.

[24] W. W. Cohen, P. Ravikumar, and S. E. Fienberg, "A comparison of string distance metrics for name-matching tasks," in *Proceedings of the IIWeb*, pp. 73–78, Acapulco, Mexico, August 2003.

[25] G. A. Miller, "Wordnet: a lexical database for English," *Communications of the ACM*, vol. 38, no. 11, pp. 39–41, 1995.

[26] G. Sidorov, A. Gelbukh, H. Gómez -Adorno et al., "Soft similarity and soft cosine measure: similarity of features in vector space model," *Computación Y Sistemas*, vol. 18, no. 3, pp. 491–504, 2014.

[27] C.-H. Chen, "A cell probe-based method for vehicle speed estimation," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E103.A-A, no. 1, pp. 265–267, 2020.

[28] C.-H. Chen, F. Song, F.-J. Hwang, and L. Wu, "A probability density function generator based on neural networks," *Physica A: Statistical Mechanics and Its Applications*, vol. 541, Article ID 123344, 2020.

[29] C.-H. Chen, F.-J. Hwang, and H.-Y. Kung, "Travel time prediction system based on data clustering for waste collection vehicles," *IEICE Transactions on Information and Systems*, vol. E102.DD, no. 7, pp. 1374–1383, 2019.

[30] R. Ichise, "Machine learning approach for ontology mapping using multiple concept similarity measures," in *Proceedings of the Seventh IEEE/ACIS International Conference on Computer and Information Science (ICIS 2008)*, pp. 340–346, IEEE, Portland, OR, USA, May 2008.

[31] S. Melnik, H. Garcia-Molina, and E. Rahm, "Similarity flooding: a versatile graph matching algorithm and its application to schema matching," in *Proceedings of the 18th International Conference on Data Engineering*, pp. 117–128, IEEE, San Jose, CA, USA, March 2002.

[32] T. K. Kim, "T test as a parametric statistic," *Korean Journal of Anesthesiology*, vol. 68, no. 6, p. 540, 2015.

[33] K. Deb, "Multi-objective genetic algorithms: problem difficulties and construction of test problems," *Evolutionary Computation*, vol. 7, no. 3, pp. 205–230, 1999.

Hindawi
Discrete Dynamics in Nature and Society
Volume 2020, Article ID 5365126, 14 pages
https://doi.org/10.1155/2020/5365126

Research Article

Fire Evacuation Process Using Both Elevators and Staircases for Aging People: Simulation Case Study on Personnel Distribution in High-Rise Nursing Home

Yameng Chen,[1] Chen Wang,[2] Jeffrey Boon Hui Yap,[3] Heng Li,[4] Hong Song Hu,[1] Chih-Cheng Chen,[5,6] and Kuei-Kuei Lai[7]

[1]*College of Civil Engineering, Huaqiao University, Xiamen 361021, China*
[2]*Construction Fujian Province Higher-Educational Engineering Research Centre, College of Civil Engineering, Huaqiao University, Xiamen 361021, China*
[3]*Department of Surveying, Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman (UTAR), Kajang 43000, Selangor, Malaysia*
[4]*Department of Building and Real Estate, The Hong Kong Polytechnic University, Hung Hom, Hong Kong*
[5]*Information and Engineering College, Jimei University, Fujian, Xiamen 361021, China*
[6]*Department of Aeronautical Engineering, Chaoyang University of Technology, Taichung, Taiwan*
[7]*Department of Business Administration, Chaoyang University of Technology, Taichung, Taiwan*

Correspondence should be addressed to Chen Wang; derekisleon@gmail.com and Chih-Cheng Chen; 3343033397@qq.com

Received 20 January 2020; Revised 2 March 2020; Accepted 11 May 2020; Published 8 June 2020

Guest Editor: Fuqiang Gu

There were increasing concerns on the possibility and suitability of using elevators for high-rise building evacuation because, through the improvement of the elevator system, the self-evacuation ability of age people is promoted as much as possible in the process of an emergency evacuation. The combined evacuation using both elevators and stairs was put into discussion. However, there was no empirical evidence and numerical simulation on emergency evacuation using both elevators and staircases for aging people in high-rise nursing homes. Therefore, using one case study, this paper simulated the emergency evacuation in a high-rise nursing home using variables such as the distribution of the elderly with different physical conditions, the proportion of the elderly in different physical conditions, the number of the elderly, the number of floors, the number of elevators used, and the priority of the elevator floor. By simulating the evacuation process in various scenarios, the general distribution strategy of high-rise nursing home and the optimal use of the elevator-stair combination during the emergency evacuation were developed. Results show that the elevator-stair combination of evacuation is more effective than using elevators or stairs alone. Increasing the number and speed of elevators can reduce evacuation time. Categorizing elderly people on each floor according to their physical conditions could reduce the evacuation time than randomly distributing them.

1. Introduction

With the advancement of science and medical technology, human beings are living longer and the world is facing the major challenge of an aging population [1–3]. Population aging has been seen in many nations [4], and most elderly people choose to live in nursing homes or hospitals during the last years of their lives [5]. More high-rise nursing homes have been built in cities to meet the needs of an aging society, but the evacuation of the elderly has become a serious problem in emergencies [6]. To reduce casualties and property losses in unexpected situations, an effective evacuation plan should be adopted [7]. Older people entering nursing homes are often seriously ill [8, 9], many of them need regular contact with the health care system to stay alive and function [10]. Those standard evacuation procedures are therefore not well suited for these vulnerable populations [11]. Aging people not only need the help of others to

evacuate safely but also hinder the evacuation of others [12]. Due to mobility problems, evacuation from the stairwell is extremely difficult for aging people [13]. Therefore, since the World Trade Center attack on September 11, there were increasing concerns about the possibility and suitability of using elevators for high-rise building evacuation [14]. Through the improvement of the elevator system, the self-evacuation ability of age people is promoted as much as possible in the process of emergency evacuation [15]. The combined evacuation using both elevators and stairs was significantly improved [16]. Computer simulation can create dynamic building information to observe the behavior of structures or people, and the design of the built environment can be promoted to respond to emergencies [17]. Agent-based evacuation simulation could simulate individual movements in high-rise building evacuation [18]. An evacuation simulation of a 40-story building found a refuge layer in the middle which could allow more people to evacuate [19]. When evacuating different age groups, crowded elderly people on the stairs were evacuated by the elevator, which effectively accelerated the evacuation [20]. A study regarding the impact of staircase design on building plane evacuation found that the evacuation time and many parameters were linearly associated [21]. However, there was no empirical evidence and numerical simulation on emergency evacuation using both elevators and staircases for aging people in high-rise nursing homes. Therefore, this study simulated the emergency evacuation in high-rise nursing homes using variables such as the use of the elevator, the distribution of the elderly with different physical conditions, the proportion of the elderly in different physical conditions, the number of the elderly, the number of floors, the number of elevators used, and the priority of the elevator floor.

2. Review Nursing Home Evacuation

Evaluating building evacuation performance design in an emergency is complex, requiring simulation tools for analysis and a large amount of manual input [22]. Built environment and evacuation behavior are decisive factors in establishing evacuation performance [23]. The built environment includes the distance from the exit, the width of the passage, the degree of congestion, and the capacity of the exit, which affect the route choices of pedestrians [24]. Human factors in evacuation include body, cognition, motivation, and social variables [25], and psychological factors affect people's behavior and panic in the case of evacuation, which makes the density of pedestrians larger and the distance between them smaller [26]. The knowledge of fire safety and emergency preparedness before an emergency has an important impact on fire response performance and evacuation response time [27, 28]. The clustering of evacuated personnel could slow down the average speed of evacuation [29]. The familiarity of the personnel with the site will greatly reduce the evacuation time, but excessive conservative or impulsive evacuation will lead to an increase in evacuation time. A small number of leaders in the pedestrian can significantly reduce evacuation time [25].

In recent years, efforts by researchers studied many high-rise buildings' evacuated simulation models. The pre-evacuation behavior was investigated using the Support Vector Machine (SVM) method in Hong Kong, and the escape route planning, safety education, equipment maintenance, and fire safety management based on BIM model buildings were established [30]. Simulation can be used to calculate the required safety exit time and available safety exit time to assess the ability to evacuate in the event of fire [31]. Applying evacuation regulations to establish a BIM-based automation system, designers and owners can check whether BIM data meets evacuation regulations, which is critical for disaster prevention systems and exit routes for tall and complex buildings [22]. The risk and route conditions were detected by a route selection model based on human organs and features, the degree of congestion of the route is determined, and the evacuation route is selected according to the personal characteristics of each occupant [32]. A stratified route planning algorithm was used to find the best evacuation path to quickly guide the evacuees to the exit [33]. A cellular automaton crowd route choice model could simulate the evacuation process of large indoor spaces in various environments [34]. When fire or chemical leak in the event of an emergency evacuation process linear programming model is viable [35], a cellular automata (CA) model was developed to describe pedestrian movement in the presence of collision avoidance during the evacuation, showing that more collision avoidance behavior hurt evacuation efficiency, but more competitive behavior had no significant positive impact on evacuation efficiency [36]. A pedestrian evacuation system for large buildings was to monitor pedestrian flows in complex facilities to assist decision makers and security agencies in emergencies [37].

The state of panic in an emergency can increase the evacuation time of people, especially elderly people and people with disabilities who are in poor health. The mental disability caused by the identification of unexpected risks in special populations has increased the average evacuation time [35]. Older people who are over 65 years old doubled the risk of dying in a fire compared to the general population [38]. Most of the elderly, living in nursing homes, need to use crutches, wheelchairs, or other occupants or firefighters to evacuate, and only a small number of elderly people can evacuate without help [39]. When these special groups are evacuated in high-rise buildings, large-scale and slow-moving elderly people such as those in wheelchairs and using a stretcher tend to block the channel, with serious ramifications for other passengers to evacuate [40–42].

Due to the special nature of high-rise building evacuation, the International Building Code allows the use of occupant evacuation elevators as a third staircase to facilitate the safe evacuation of high-rise building personnel [43]. As a result, high-rise buildings are equipped with elevator evacuation that can be used for the elderly and disabled in unexpected situations [44–46]. There is an upper limit to the optimization process of elevator-assisted evacuation [47]. The elevator evacuation time is determined by influencing

factors such as the number of elevator evacuation, evacuation floor height, elevator speed and acceleration, elevator capacity, and elevator door width [28]. The spacing design of the evacuation floor is directly related to the characteristics of the elevator and the occupants of the building. The evacuation process can be optimized while the appropriate proportion of building occupants are transported to the ground by elevators while others are evacuated by stairs [24]. Using elevators to move all passengers to the ground safety point is not the best solution [29]. People who use stairs or elevators to evacuate are mainly affected by the floor. In an emergency, the waiting time of the passengers and the proportion of waiting people are affected by the height of the floor, and it is unreasonable to wait for the elevator indefinitely [48]. Even if people choose elevator evacuation, they may not wait if it takes a long time to reach [49]. Stair evacuation plays a vital role in building evacuation, as the evacuation time can be predicted by a simulation model and architects can develop evacuation plans and strategies based on simulation results [50]. The consolidation behavior of stairwells could reduce the speed of pedestrian flow. The stairs are the major exporters of high-rise buildings [51, 52]. As the population grows, the impact of obstruction caused by people with disabilities on evacuation time becomes more apparent [53, 54].

3. Research Methods and Procedures

The basic model used in this simulation was a 17-story nursing home for aging people located in Fujian, China. Figure 1 shows the layout of the first and third floors of the nursing home in Pathfinder simulation software. It is 72 m long, 15 m wide, and 61.6 m high. The first floor is the lobby office area, and the second floor includes the chess room, the dining hall, and the activity room. The third floor is the residential area for the elderly. The first floor has a height at 4 m and the 2nd to 17th floors have a height of 3.6 m each. Each floor has 56 beds, and the 4th to 17th floors were designed as same as the third floor. The building has three exits, two stairs with a tread of 0.3 m and a riser height of 0.15 m, and two elevators with a length of 3.6 m and a width of 2.7 m. The following is the setting of elevator parameters in Pathfinder simulation software. The elevator load was calculated according to the software according to the size of the personnel and the size of the elevator in STEERING mode. The nominal load of the two elevators in this model was 29 people. The open and close time of the elevator doors was calculated by 7 s, and the acceleration of the elevator was 1.2 m/s^2. The elevator bounds were from the 1st floor to the 17th floor, and the first floor was the discharge layer of the personnel. In the event of an emergency, the person connected to the elevator on the floor should be sent directly to the first floor for evacuation. In this process, even if the elevator did not reach the nominal load, it would not pick up more people on other floors. The modeling of types of people living in this nursing home was complicated. According to the physical conditions of the elderly, the elderly people could be divided into the no-aid elderly people (NAEP) and elderly people who use auxiliary tools. Old people who use assistive tools were divided into the elderly people who use single-crutch (SCEP), the elderly people who use double-crutches (DCEP), the elderly people who use a manual wheelchair (MWEP), the elderly people who use power wheelchair (PWEP), and the elderly people who use Evac + Chair (ECEP). Evac + Chair evacuate the elderly with a poor physical condition, which is 50% faster than other devices, so the elderly using stretchers choose to use Evac + Chair for evacuation. The formula for calculating the total evacuation time of personnel is as follows:

$$T = T_0 + T_1, \tag{1}$$

where T is the total evacuation time, T_0 is the evacuation exercise time, and T_1 is the evacuation delay time. The parameter setting and evacuation delay time in the simulation were set manually. Tables 1 and 2 show the parameter values of different types of elderly people in Pathfinder software settings. According to the experimental investigations [55–61], the movement speed, the width of the occupation, the separation distance from other people, and the delay time of evacuation were obtained in Table 1. For example, Speed of Nurse in the software is 1.549 m/s, Shoulder Width is 40 cm, Comfort Distance is 0.08 m, and Delay Time is 0.8 s. The speed of NAEP in the software is 1.274 m/s, Shoulder Width is 40 cm, Comfort Distance is 0.08 m, and Delay Time is 0.8 s. Similarly, parameter values of other types of elderly people can be set. The width of the occupation is the projection width of the elderly on the ground when they are stationary. For example, the width of the occupation of a standing elderly person is shoulder width, and the width of the occupation of an elderly person in a wheelchair is the width of a wheelchair. The comfortable distance is the distance that people habitually maintain with each other. Within this distance, personnel can switch directions freely and avoid collisions. They must also follow each other and use space reasonably. For example, the auxiliary tools used by MWEP, PWEP, and ECEP have a large width, and any movement or rotation requires a large space. To ensure normal movement and avoid collisions, the required comfort distance is large. When an emergency occurs, the distance between the elderly will be less than the comfortable distance, and even events such as crowding and stamping will occur. Therefore, the comfort distance of different types of elderly in the emergency simulated in this paper is the minimum distance kept between people, which is 0.08 m. The delay time is the difference between when an emergency occurs and when people evacuate. The delay time for a self-evacuated person is the person's reaction time. It is known from the literature that the average human response time is 0.8 s. ECEP needs to wait for the help of others before the evacuation, so the delay time of ECEP is the time to assist people to reach their location. The speed of the upstairs and the horizontal movement speed of the assisting staff

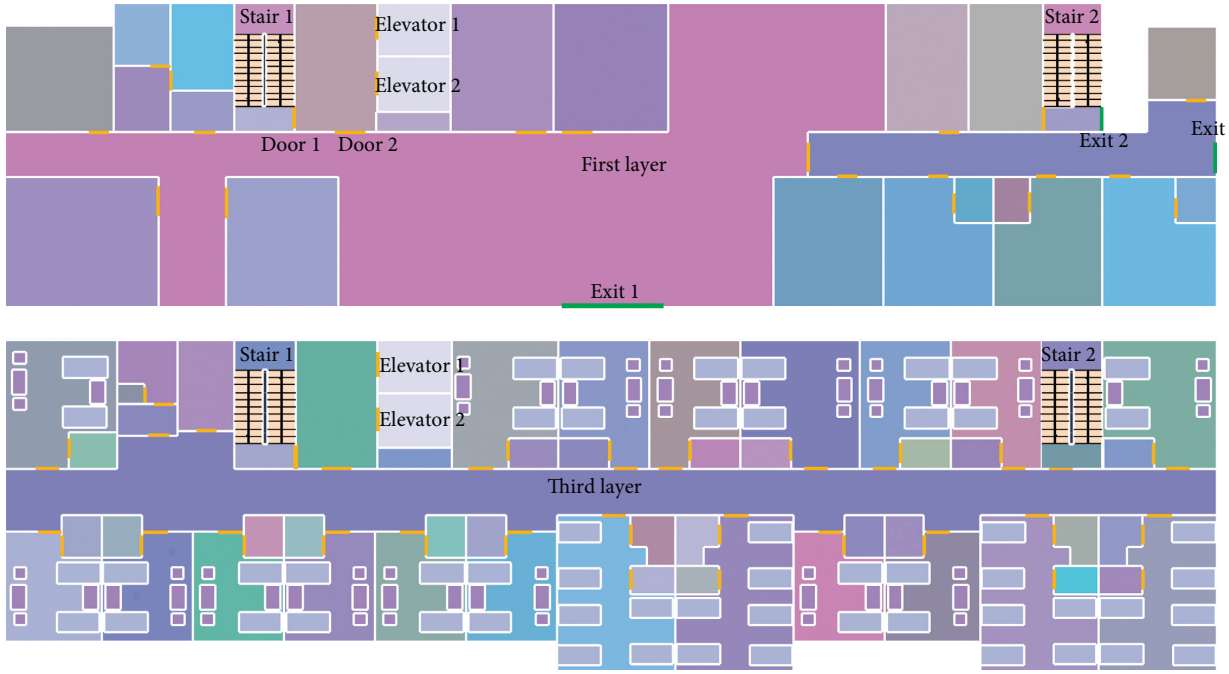


Figure 1: Floor plan of a high-rise nursing home for aging people.

Table 1: Attributes of the elderly with different physical conditions.

Types	Moving speed (m/s)		Occupied width (m)	Comfort distance (m)	Delay (s)
	Horizontal	Stair-ascending			
Nurse	1.549	1.146	0.4	0.08	0.8
NAEP	1.274	0.85	0.4	0.08	0.8
SCEP	0.873	0.433	0.45	0.13	0.8
DCEP	0.779	0.332	0.5	0.18	0.8
MWEP	0.64	0	0.98	0.87	0.8
PWEP	0.7	0	0.98	0.87	0.8
ECEP	1.5	0.81	1.1	1.2	See Table 2

Table 2: Delay time for elderly people using ECEP on different floors.

Floors	Delay time (s)	Floors	Delay time (s)
2	14.73	10	46.06
3	18.64	11	49.98
4	22.56	12	53.90
5	26.48	13	57.82
6	30.39	14	61.73
7	34.31	15	65.65
8	38.23	16	69.57
9	42.15	17	73.49

were available according to the survey value and the time formula for assisting personnel to reach the floor is as follows:

$$T_3 = \frac{H}{V_0} + \frac{L}{V_1}, \tag{2}$$

where T_3 is the time to assist the person to reach the floor of the old men, H is the height of the floor where the old men were located, L is the horizontal distance of an old man's position from the stairs, V_0 is the speed to assist the staff to go upstairs, and V_1 is the horizontal speed of the assisting personnel. Delay time for elderly people using ECEP on different floors is listed in Table 2.

4. Data Interpretation and Analysis

4.1. Simulation Using Elevator Evacuation. Most of the elderly people living in nursing homes were unable to take care of themselves. Old people with poor physical condition need the help of the medical staff and daily care, so it is necessary to arrange nursing staff for the floor where the elderly who are in poor health live. The number of elderly people in the nursing home in Pathfinder software was set in Table 3. In the event of an emergency, elderly people in wheelchairs need to replace the evacuated aids with crutches if they need to evacuate through the stairs. Table 4 shows five different scenarios in Pathfinder software simulation, including (i) all elevators, (ii) all stairs, (iii) one elevator + two stairs, (iv) two elevators + two stairs, and (v) when the maximum speed of the elevator increased. The elderly in the nursing home were randomly distributed according to the proportion of the people in Table 3. The priority of the

Table 3: Number and distribution of the elderly using different auxiliary tools.

Items	Nurse	NAEP	SCEP	DCEP	MWEP	PWEP	ECEP	Total
Occupants	24	56	112	108	108	108	324	840
Proportion (%)	2.86	6.67	13.33	12.86	12.86	12.86	38.57	100.00

Table 4: The description of Scenarios A1–A5.

Scenario	Description
A1	All the elderly are evacuated by elevator; the elevator speed is 2.5 m/s
A2	All the elderly use stairs to evacuate
A3	All the elderly are evacuated by an elevator + two stairs; the elevator speed is 2.5 m/s
A4	All the elderly are evacuated by two elevators + two stairs; the elevator speed is 2.5 m/s
A5	All the elderly are evacuated by two elevators + two stairs; the elevator speed is 5 m/s

elevator was from the upper to the lower floors. To eliminate the chance factor, each case was simulated 10 times. Table 5 recorded the occupant evacuation time of ten simulations in Scenarios A1–A5.

The evacuation time using only elevators and only stairs reached the maximum. The evacuation time using only elevators and only stairs were 52% and 30%, respectively, higher than using both elevators and stairs. The combination of using both the stairs and elevators could shorten the evacuation time. The percentage of evacuation time that no elevator has increased over one elevator was (2778–2354)/2778 = 15.3%. The percentage of evacuation time increased using one elevator compared to that using two elevators was (2354–2131)/2354 = 9.5%. From Scenario A4 and Scenario A5, the speed of the elevator also played an impact on the evacuation time, and the percentage of evacuation time reduction when the speed of the elevator increased from 2.5 m/s to 5 m/s was calculated as (2131–1809)/2131 = 15.1%. Increasing the speed and number of elevators resulted in a significant reduction in evacuation time. When elevators stairs were used for evacuation, the increased number of elevators did not increase the elevation speed; thus, in the case of an emergency evacuation, the elevator speed should be expedited.

Different floors and different types of seniors have different delay times. The delay time values are shown in Tables 1 and 2. In pathfinder simulation software, adding "Wait" in "Behavior" means that the delay time for older people is different. Although the use of a combination of elevators and stairs to evacuate and increase the speed of the elevator can speed up the evacuation efficiency, the random distribution of different types of elderly in the building has caused a large congestion phenomenon. For example, ECEP is randomly distributed on each floor, and the ECEP on the top floor waits for people to assist for a long time. Figure 2 is a graph showing the cumulative number of evacuations per door over time in Scenario A5. The time for completing the evacuation of stairs 1, stairs 2, and elevators are quite different. The vast majority of elderly people use elevators to evacuate, and the stairs cannot perform the evacuation function, extending the overall evacuation time.

4.2. Influence of Different Types of Personnel Distribution on Evacuation Time. The distribution of the elderly with different physical conditions could have an influence on evacuation time, and the scene descriptions of three different distributions of the elderly in Pathfinder software were listed in Table 6. Combining elevators and stairs and increasing the maximum speed of the elevator could reduce the evacuation time; thus, the following simulation adopted the evacuation mode of elevator stairs combination, and the maximum running speed of the elevator was 5 m/s. Table 7 shows the number of distribution of personnel on each floor in Scenario B2 in Pathfinder software. There numbers 4-5 in Table 7 represent the number of elderly people from the fourth floor to the fifth floor, and the number of elderly people living on the fourth floor and the fifth floor was 56. Table 8 shows the distribution of personnel on each floor in Scenario B3 in Pathfinder software. The proportion of each type of old people in Scenarios B1, B2, and B3 was the same, and the distribution of the area in which the elderly people were staying was different. All the elderly people randomly chose the evacuation methods, and the priority of the elevator was from the upper to the lower.

According to the different personnel distributions in Scenarios B1, B2, and B3, 10 simulations were performed, respectively, and the simulation results were shown in Figure 3.

Scenario B1 used a longer evacuation time than Scenario B2 because different types of elderly people were randomly assigned to each floor, and elderly people with poor physical condition could easily cause congestion on the stairs during evacuation through stairs. If the ECEP was on the middle floor, when the assisting person arrived at the target floor to help evacuate, the person on the upper floor was evacuated to that floor. If ECEP chose the stairs to evacuate at this time, the space occupied by the evacuation was the largest, and the faster moving speed of the assisting personnel could not function, so the evacuation efficiency was very low. When this happened, not only the ECEP wads are congested in the stairs but also the old people who used crutches or even self-care could not reach the bottom layer due to the blockage of the stairs, so the evacuation time was the longest. The ECEP was uniformly arranged on the top layer. When the assisting personnel reached the top level to assist the ECEP

Table 5: Occupant evacuation time of ten simulations in Scenarios A1–A5 (unit: s).

Scenario	1	2	3	4	5	6	7	8	9	10	Mean	Percent
A1	3209	3260	3098	3367	3298	3198	3321	3048	3306	3386	3249	52
A2	2725	2743	2681	2801	2825	2722	2941	2595	2856	2895	2778	30
A3	2306	2340	2302	2384	2360	2306	2503	2165	2406	2463	2354	10
A4	2091	2160	2013	2201	2156	2101	2305	1864	2150	2264	2131	0
A5	1780	1670	1722	1826	1876	1798	1984	1657	1874	1905	1809	−15

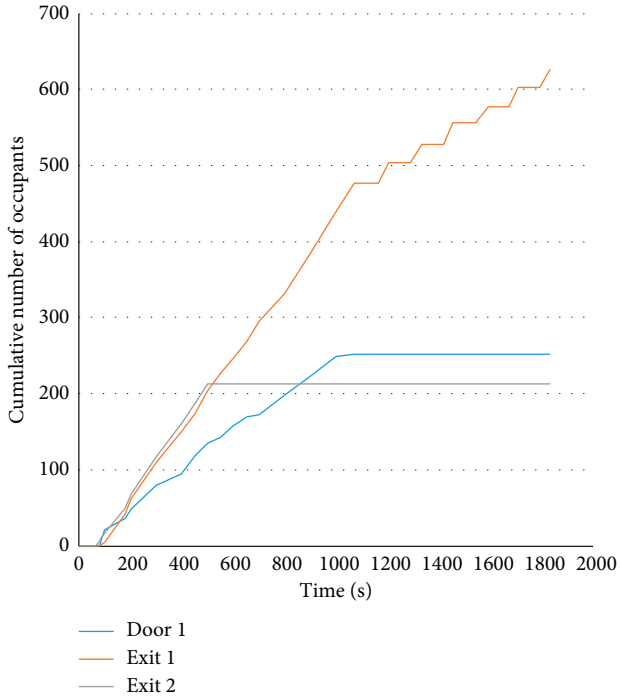


Figure 2: The cumulative number of evacuations per door over time in Scenario A5.

evacuation, the underlying NAEP was evacuated in an orderly manner, and the evacuation efficiency was improved. The evacuation time of Scenario B3 was smaller than that of Scenario B2 because the time required to assist the person to reach the bottom layer was shorter. Although ECEP occupied a large area when evacuated, the evacuation rate was faster than other elderly people. When the assisting personnel reached the ground floor, they could quickly evacuate through the stairs. Meanwhile, the seniors of the upper level did not reach the lower level yet, and the old people could be evacuated through the stairs as quickly and orderly as possible. When the old man who evacuated through the stairs reached the lower level, the lowest number of layers of the ECEP evacuated through the stairs was evacuated, so the congestion was not serious. The ECEP of the top floor in Scenario B2, no matter being evacuated by stairs or elevator, needed assistance from personnel to reach the stairs or elevators. It took longer for the personnel to reach the top floor thus the ECEP delay time was longer. The reverse movement of assisting personnel and evacuated old people in the stairwell affected the overall evacuation efficiency, making the stairs congested and prolonging the evacuation time. The evacuation time of Scenario B2 and Scenario B3 was shorter than that of Scenario B1, indicating that the same type of elderly people should be arranged on the same floor, which not only facilitated the care and help of medical personnel but also reduced the evacuation time in an emergency. Scenario B3 was shorter than the evacuation time used by Scenario B2, indicating that ECEP should be placed at the lowest level, and evacuation time could be shortened when an emergency occurred.

4.3. Impact of Property and Quantity of Elderly on Evacuation Time. ECEP was arranged at the bottom layer, and NAEP was arranged at the top layer, which effectively shortened the evacuation time. Therefore, the following model assumed that ECEP was at the bottom and NAEP was at the top. When the proportion of the elderly with different physical conditions was different, the evacuation time was also affected. The situation simulated above was the largest proportion of ECEP while the following assumed that the proportion of different types of elderly people was just the opposite. Table 9 shows that ECEP had the smallest number of people and NAEP had the largest number in Pathfinder software. Table 10 shows the distribution of the number of elderly people on each floor in Scenario C2 in Pathfinder software. The change in the number of people on each floor had an impact on the evacuation time. Scenario C3 was to reduce the number of people on each floor from 56 to 30, but the proportion of each type of old people and the living floor was the same as Scenario C1.

Figure 4 presents the simulation results of Scenarios C1, C2, and C3 when elevators were evacuated on different floors. The numbers 3–17 in the table referred to the use of elevators to evacuate from the 3rd floor to the 17th floor and the use of stairs for evacuation on other floors. The numbers 4–17 referred to the 4th floor to the 17th floor using elevator evacuation and other floors using stairs to evacuate. The number 0 means that the elderly on all floors were evacuated by stairs and no floors were evacuated by elevators. In this simulation, the priority of the elevator was from upper to lower floors.

The evacuation time used by Scenario C1 was always the longest. Scenario C2 had the fastest reduction in evacuation time when the number of layers used in the elevator was reduced. When the number of layers used in the elevator was on the 11th–17th floors, Scenario C2 reached the shortest evacuation time at 719 s. The evacuation time of Scenario C1 and Scenario C3 was similarly reduced, and both scenarios had the shortest evacuation time on the 13th–17th floors, and then the evacuation time continued to increase as the elevator used the number of layers. The simulation results

Table 6: Description of Scenarios B1, B2, and B3.

Scenario	Description
B1	Same as A5
B2	The old man living on the upper floor is from ECEP to NAEP
B3	The old man living on the lower floor is from NAEP to ECEP

Table 7: Number of different types of elderly people on each floor in Scenario B2.

Floor	Nurse	NAEP	SCEP	DCEP	MWEP	PWEP	ECEP	Total
3		56						56
4-5			56					56
6-7	2			54				56
8-9	2				54			56
10-11	2					54		56
12–17	2						54	56
Total	24	56	112	108	108	108	324	840
Proportion (%)	2.86	6.67	13.33	12.86	12.86	12.86	38.57	100.00

Table 8: Number of different types of elderly people on each floor in Scenario B3.

Floor	Nurse	NAEP	SCEP	DCEP	MWEP	PWEP	Evac + chair	Total
3–8	2						54	56
9-10	2					54		56
11-12	2				54			56
13-14	2			54				56
15-16			56					56
17		56						56
Total	24	56	112	108	108	108	324	840
Proportion (%)	2.86	6.67	13.33	12.86	12.86	12.86	38.57	100.00

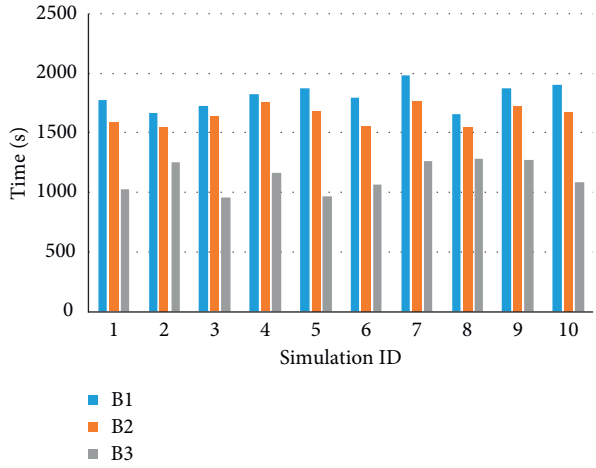


Figure 3: Occupant evacuation time of ten simulations in Scenarios B1, B2, and B3.

also demonstrated the need for a reasonable combination of elevator and stair evacuation. When the final evacuation time interval between the elevator and the stairs was small, the minimum time for evacuation was reached. Excessive use of elevators or stair evacuation could result in extended evacuation time. The longest evacuation time of the above three scenarios occurred when the number of elevators used was on the 3rd–17th floors, indicating that the use of elevators to evacuate all the elderly were the most unreasonable evacuation method. The shortest evacuation time of Scenarios C1, C2, and C3 were 852 s, 719 s, and 539 s, respectively, indicating that the number of elderly people with different physical conditions could affect the evacuation time. When the proportion of the elderly with the poor physical condition was high, the evacuation time increased. When the total number of elderly people living decreased, the evacuation time reduced.

4.4. Influence of Different Living Floors in High-Rise Nursing Home on Evacuation Time. Since most of the elderly people living in nursing homes were in poor physical condition needing the care from medical staff, the simulator set the largest proportion of ECEP and the smallest proportion of NAEP. Table 11 lists the description of different scene settings in Pathfinder software. Table 12 shows the representations when the number of elevators used in different scenarios was different. For instance, Case 1 indicated that Scenarios D1, D2, and D3 were evacuated by elevators on floors 6–13, 9–17, and 16–27, and stairs were used for evacuation on other floors. Case 2 indicated that Scenarios D1, D2, and D3 were evacuated by elevators on floors 7–13, 10–17, and 17–27 and evacuated by stairs on other floors. Figure 5 shows the evacuation time for each of the three scenarios. Figure 6 illustrates the relationship between the number of elderly people in nursing homes in Scenarios C1,

Table 9: Description of Scenarios B1, B2, and B3.

Scenario	Description
C1	Same as B3
C2	Number of people per floor is 56; the proportion of people is shown in Table 11
C3	Number of people per floor is 30; the proportion of people is shown in Table 8

Table 10: Number of different types of elderly people on each floor in Scenario C2.

Floor	Nurse	NAEP	SCEP	DCEP	MWEP	PWEP	ECEP	Total
3	2						54	56
4-5	2					54		56
6-7	2				54			56
8-9	2			54				56
10-11			56					56
12–17		56						56
Total	14	336	112	108	108	108	54	840
Proportion (%)	1.67	40.00	13.33	12.86	12.86	12.86	6.43	100.00

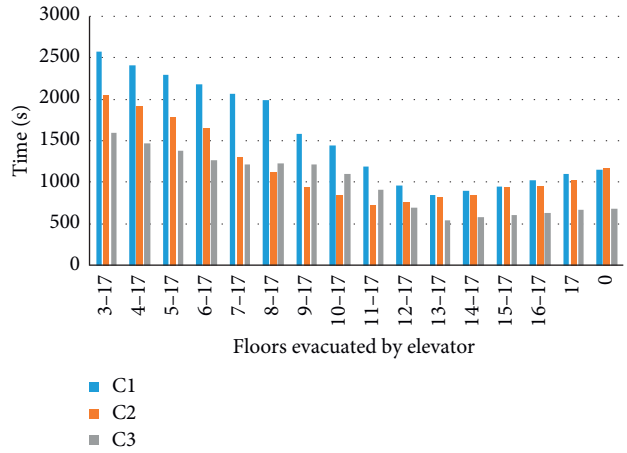


Figure 4: Occupant evacuation time of ten simulations in Scenarios C1, C2, and C3.

C2, C3, and D1 with the shortest evacuation time. The shortest evacuation time in Scenarios C1, C2, C3, and D1 were indicated when the number of layers used in the elevator was 13–17, 11–17, 13–17, and 10–13.

Figure 5 shows that the shortest evacuation time was proportional to the number of floor levels. As the number of floor levels increased, the minimum evacuation time increased. When the total number of floors in the nursing home building was 13, 17, and 27, the optimal number of floors for the elevator was floored 10–13, 13–17, and 20–27; the other floors were evacuated by stairs. Figure 6 illustrates that the trends of the number of elderly people in Scenarios C1, C3, and D1 were similar over time. When the proportion of different types of elderly people was the same, the evacuation efficiency of the elderly was also similar, regardless of the number of people on each floor and the number of floors. Scenario C2 had a slower evacuation speed in the first 400 s than Scenario C1 and quickly exceeded Scenario C1 after 400 s. During this period, the elevator was from the upper to the lower floors. The elderly on the upper floor of Scenario C2 was in good physical condition. Therefore, Scenario C2 was evacuated by the elevator in a short time, and the number of people was more than that of Scenario C1. However, the total number of people evacuated in the first 400 s of Scenario C2 was less than that of Scenario C1, indicating that the number of evacuated people in Scenario C2 was less than that in Scenario C1. The ECEP of the third layer of the first 400 s in Scenario C2 needed to wait for the assisting personnel to arrive. When the assisting personnel arrived, the personnel on the upper floor were evacuated to the third floor, and it was difficult for ECEP to enter the stairs. In Scenario C1, since the lower floors were all ECEP, other types of people in the upper layer did not arrive when the ECEP of the third layer was evacuated using stairs. Therefore, the ECEP of the following floors could be evacuated in an orderly manner. Although ECEP required a large delay time and a large footprint, the evacuation speed of ECEP was faster than that of other elderly people with the help of assisting personnel. Therefore, the number of the evacuation of Scenario C1 in the first 400 s was lower than that of Scenario C2.

4.5. Number of Floors on Optimal Elevator Use and Priority of Elevator Floor. The proportions of different types of old people in Scenarios C1, C3, D1, and D3 are the same. According to the analysis, the percentage of the total number of people using elevator evacuation in the shortest evacuation time in the four scenarios is shown in Table 13.

Percentage 1 in Table 13 refers to the percentage of the number of floors that were evacuated by the elevator to the total number of floors, and Percentage 2 refers to the percentage of the number of people who used the elevator evacuation to the total number of people. When the priority of the elevator-mounted floor was from the upper layer to the lower layer, the values of Percentages 1 and 2 of Scenarios C1, C3, D1, and D3 were identical as 29% and 33%, respectively. Scenario C2 had a different proportion of elderly people with different physical conditions. The percentage of people who used elevators to evacuate was 47%, which was different from the other four scenarios. When the

Table 11: The description of Scenarios D1, D2, and D3.

Scenario	Description
D1	The number of floors is 13 and the number of people on each floor is 56; the proportion of people is according to Table 8
D2	Same as C1
D3	The number of floors is 27 and the number of people on each floor is 56; the proportion of people is according to Table 8

Table 12: Representation of floors using elevators in different scenarios.

Scenario	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 9
D1	6–13	7–13	8–13	9–13	10–13	11–13	12–13	13	0
D2	9–17	10–17	11–17	12–17	13–17	14–17	15–17	16–17	17
D3	16–27	17–27	18–27	19–27	20–27	21–27	22–27	23–27	24–27

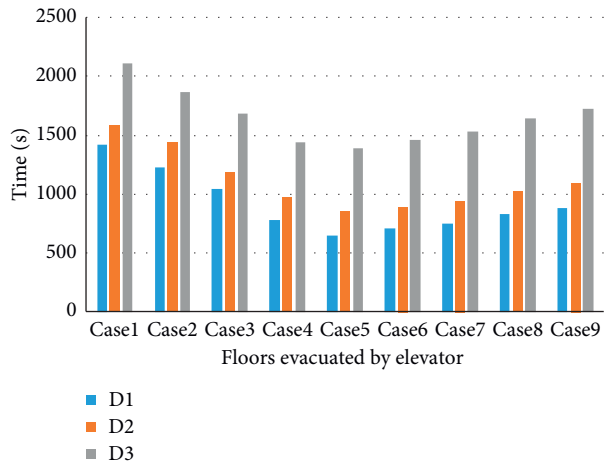


Figure 5: Occupant evacuation time of ten simulations in Scenarios D1, D2, and D3.

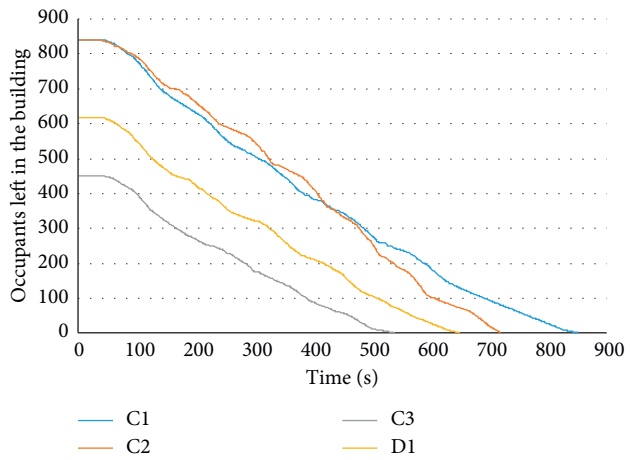


Figure 6: Curve of the number of elderly people in nursing homes over time.

proportion of the elderly with different physical conditions in nursing homes was the same, the proportion of the number of layers used by the elevators to the total number of floors and the number of people using the elevators was the total when the minimum evacuation time was reached. In emergencies, the optimal number of staying floors of the elevator in one nursing home could be set referring to other nursing homes with similar proportions, which could make the evacuation method efficient. Figure 7 shows the cumulative number of evacuations per door over time in Scenarios C1, C2, C3, D1, and D3 in the case of Table 13. Stairs 1 and 2 have the same size, and the elderly can choose the stairs to evacuate reasonably. Door 1 is an evacuation door for Stair 1, and Exit 2 is an evacuation door for Stair 2. The two curves are very close, indicating that the evacuation efficiency and cumulative evacuation times of the two stairs are almost the same. The difference between the cumulative evacuation number of Exit 1 and Door 1 is the cumulative evacuation number of elevators. Although the evacuation efficiency and cumulative evacuation number of elevators are higher than those of stairs, the cumulative evacuation time is the same as that of stairs. Stairs and elevators can maintain continuous evacuation throughout the evacuation process, achieving the best overall evacuation effect.

The priority of the elevator floor could affect the evacuation of the elderly. The above analysis assumed that the priority of the elevator floor was from the upper to the lower. Three scenarios were set in Table 14 to analyze the situation in Pathfinder software when the elevating floor priority of the elevator was from the lower layer to the upper layer. Figure 8 illustrates the evacuation time when the number of elevators was different in Scenarios E1, E2, and E3. Figure 9 shows the shortest evacuation time for Scenarios C1, C2, C3, E1, E2, and E3.

Figure 8 shows that when the priority order of the elevator was from the lower layer to the upper layer and the minimum evacuation time was reached, the number of elevators used in Scenarios E1, E2, and E3 was floors 3–5, 3–4, and 3–7, respectively, and the number of elevators used in Scenarios C1, C2, and C3 was floors 13–17, 11–17, and 13–17, respectively. In Scenario E, the underlying ECEP needed assistance from the facilitators; thus, the elderly needed to wait for the arrival of the facilitators in place, resulting in a delay. Because ECEP occupied a large position, it was very inflexible in the process of moving horizontally to the elevator. The movement into the elevator was slow, affecting the number of times the elevator was mounted. The large area of the ECEP caused the elevator to carry a small number of people per raft, which resulted in a longer time required for each floor of the elevator. Figure 9 shows that

Table 13: Relationship between the number of evacuated elevator evacuees and the total number of evacuees in the shortest evacuation time under Scenarios C1, C3, D1, and D3.

Scenario	Shortest evacuation time	Floors	Number of floors	Total number of floors	Number of elevator users	Total evacuees	Percentage 1	Percentage 2
C1	852	13–17	5	17	280	840	29	33
C2	719	11–17	7	17	392	840	41	47
C3	539	13–17	5	17	150	450	29	33
D1	650	10–13	4	13	224	616	31	36
D3	1391	20–27	8	28	448	1400	29	32

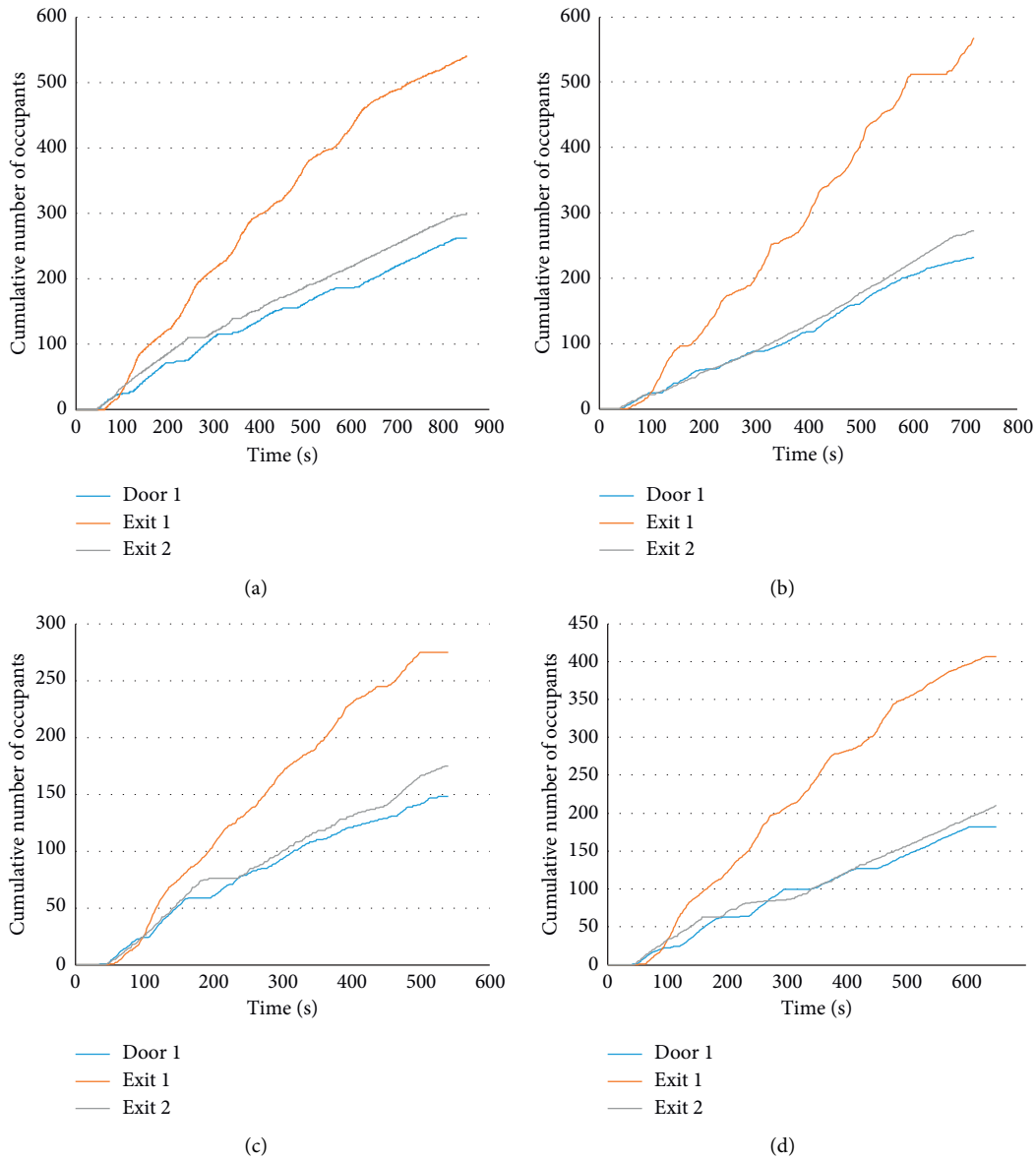
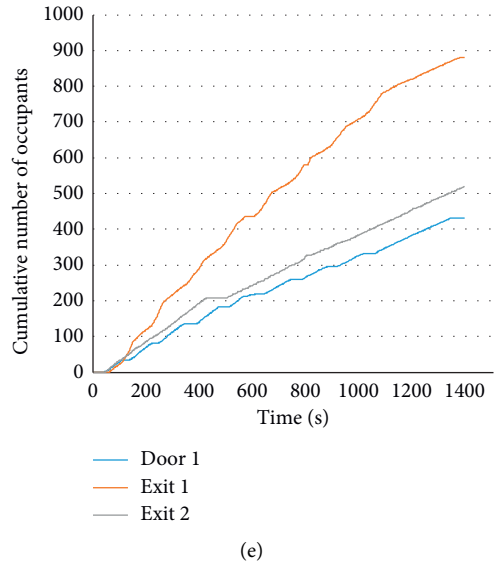


Figure 7: Continued.



(e)

Figure 7: The cumulative number of evacuations per door over time in Scenarios (a) C1, (b) C2, (c) C3, (d) D1, and (e) D3 in the case of Table 13.

Table 14: Description on Scenarios E1, E2, and E3.

Scenario	Description
E1	The number of people on each floor is 56; the proportion of the people is shown in Table 8
E2	The number of people on each floor is 56; the proportion of the people is shown in Table 11
E3	The number of people on each floor is 30; the proportion of the people is shown in Table 8

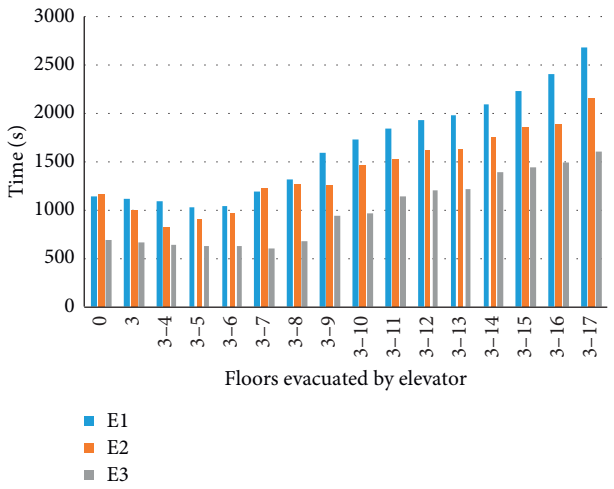


Figure 8: Occupant evacuation time of ten simulations in Scenarios E1, E2, and E3.

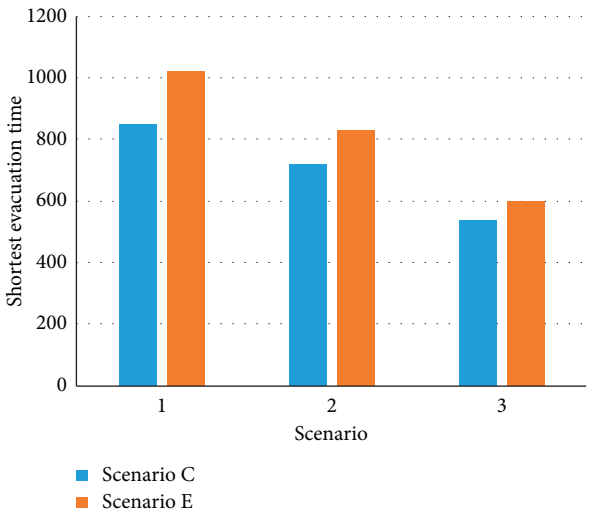


Figure 9: Minimum evacuation time for Scenarios C1, C2, C3, E1, E2, and E3.

the shortest evacuation time used in Scenarios C1, C2, and C3 was smaller than that of Scenarios E1, E2, and E3. Therefore, the priority order of the elevators was shorter from the upper level to the lower level than that from the lower level to the upper level, regardless of the physical condition of the occupants or the number of occupants on each floor.

4.6. Cumulative Number of Evacuations per Gate for Scenario C1 at the Shortest Evacuation Time. Figure 10 presents the cumulative evacuation number of each door as a function of time when the minimum evacuation time was reached and elevator evacuation was used when the number of floors used for elevator evacuation was 13–17 in Scenario C1. Figure 10 shows that all elderly people evacuated from Door

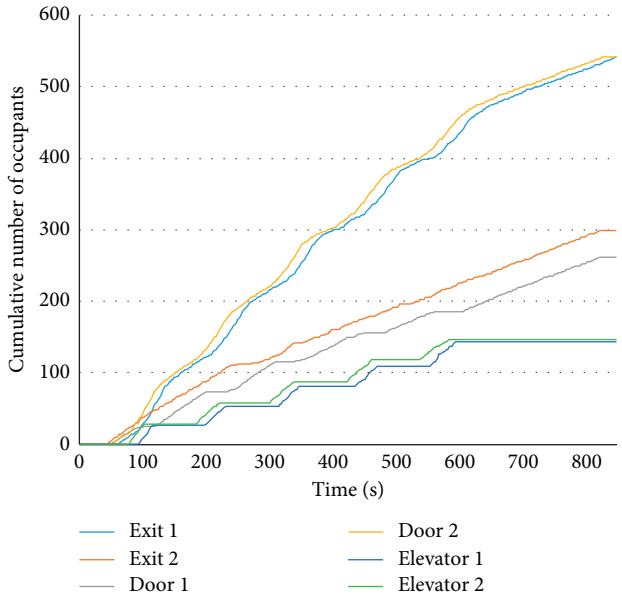


Figure 10: The cumulative number of evacuations per door over time.

2 entered Exit 1. The use of Stair 1 and elevators all reached Exit 1 through Door 1, causing great congestion in the elevator front room. Old people in the elevator could not get out of the elevator in time when they reached the ground floor. People evacuated through the stairs were too crowded to be evacuated to the exit in time. Therefore, the space in the front room of the elevator and the width of Door 2 should be increased. The trend of the evacuation curves of Exit 2 and Door 2 was similar, indicating that the evacuation efficiencies of the two stairs were similar. Elevator 1 and Elevator 2 had the same cumulative evacuation curve. At 600 s, the elevator was completed first, and then only two stairs were evacuated. The evacuation time through the two stairs was almost the same. Stair 1 had fewer evacuations than Stair 2 because Stair 1 needed to share the exit with the elevator when evacuating to the ground exit. If the exit was overcrowded, personnel would choose the best exit to avoid congestion. Therefore, the persons who originally chose Stair 1 would also change the evacuation route and chose Stair 2. According to the above analysis, it is necessary to increase the space of the common area in the elevator and the stairs and the width of the evacuation passage, to reduce the congestion and the evacuation time.

5. Conclusions

This study simulated the evacuation in a typical high-rise nursing home in several scenarios including the distribution of the elderly with different physical conditions, the proportion of the elderly in different physical conditions, the number of the elderly, the number of floors, the number of elevators used, and the priority of the elevator floor. By simulating the evacuation process in these scenarios, the general distribution strategy of high-rise nursing home and the optimal use of the elevator-stair combination during the emergency evacuation were developed. Results show that the elevator-stair combination of evacuation is more effective than using elevators or stairs alone. Increasing the number and speed of elevators can reduce evacuation time. Categorizing elderly people on each floor according to their physical conditions could reduce the evacuation time than randomly distributing them. Elderly people with better physical conditions are advised to be arranged on upper floors and those with worse physical conditions are to be arranged on lower floors to shorten the evacuation time. Although ECEP needs to wait for assistance personnel to arrive and takes up more space during the evacuation, it moves faster with the help of assistance personnel than other types of elderly people. Pacing ECEP on high floors could generate a reverse flow, causing the stairs to be blocked. The evacuation efficiency of elderly people was not affected by changing the total number of floors. When the proportion of different types of elderly people was fixed, the trend of the relationship between evacuation number and time was consistent, regardless of the number of floors and the number of people on each floor. The longest evacuation time was used when the elevator started from the lower layer to the upper layer, and the shortest evacuation time was used when the elevator started from the upper layer to the lower layer. In the future simulation, modern facilities such as life slides could be considered to combine with stairs and elevators for emergency evacuation.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

Chen Wang conceptualized the study; Yameng Chen carried out the methodology; Jeffrey Boon Hui Yap prepared and wrote the original draft; Heng Li validated the study; Hong Song Hu helped with the software; Chih-Cheng Chen validated and wrote the article, review and editing; and Kuei-Kuei Lai wrote the article.

Acknowledgments

This research was funded by Huaqiao University, Grant no. 17BS201, and Quanzhou City Government, Grant no. 600005-Z17X0234. The APC was funded by Huaqiao University.

References

[1] L. Dueñas, M. Balasch i Bernat, S. Mena del Horno, M. Aguilar-Rodríguez, and E. Alcántara, "Development of predictive models for the estimation of the probability of suffering fear of falling and other fall risk factors based on posturography parameters in community-dwelling older

adults," *International Journal of Industrial Ergonomics*, vol. 54, pp. 131–138, 2016.

[2] W.-Y. Yu, H.-F. Hwang, M.-H. Hu, C.-Y. Chen, and M.-R. Lin, "Effects of fall injury type and discharge placement on mortality, hospitalization, falls, and ADL changes among older people in Taiwan," *Accident Analysis & Prevention*, vol. 50, pp. 887–894, 2013.

[3] S. D. Choi, L. Guo, D. Kang, and S. Xiong, "Exergame technology and interactive interventions for elderly fall prevention: a systematic literature review," *Applied Ergonomics*, vol. 65, pp. 570–581, 2017.

[4] E. Mitty and S. Flores, "Aging in place and negotiated risk agreements," *Geriatric Nursing*, vol. 29, no. 2, pp. 94–101, 2008.

[5] T. Reyniers, L. Deliens, H. R. Pasman et al., "International variation in place of death of older people who died from dementia in 14 European and non-European countries," *Journal of the American Medical Directors Association*, vol. 16, no. 2, pp. 165–171, 2015.

[6] J. Koo, Y. S. Kim, and B.-I. Kim, "Estimating the impact of residents with disabilities on the evacuation in a high-rise building: a simulation study," *Simulation Modelling Practice and Theory*, vol. 24, pp. 71–83, 2012.

[7] E. Pourrahmani, M. R. Delavar, and M. A. Mostafavi, "Optimization of an evacuation plan with uncertain demands using fuzzy credibility theory and genetic algorithm," *International Journal of Disaster Risk Reduction*, vol. 14, pp. 357–372, 2015.

[8] Y. Young, J. Kalamaras, L. Kelly, D. Hornick, and R. Yucel, "Is aging in place delaying nursing home admission?" *Journal of the American Medical Directors Association*, vol. 16, no. 10, pp. 900-901, 2015.

[9] C. A. Estabrooks, M. Hoben, J. W. Poss et al., "Dying in a nursing home: treatable symptom burden and its link to modifiable features of work context," *Journal of the American Medical Directors Association*, vol. 16, no. 6, pp. 515–520, 2015.

[10] L. Saarnio, A.-M. Boström, R. Hedman, P. Gustavsson, and J. Öhlén, "Enabling at-homeness for residents living in a nursing home: reflected experience of nursing home staff," *Journal of Aging Studies*, vol. 43, pp. 40–45, 2017.

[11] S. C. Blake, D. H. Howard, H. Eiring, and S. Tarde, "San Diego's area coordinator system: a disaster preparedness model for US nursing homes," *Disaster Medicine and Public Health Preparedness*, vol. 6, no. 4, pp. 424–427, 2012.

[12] M. Willoughby, C. Kipsaina, N. Ferrah et al., "Mortality in nursing homes following emergency evacuation: a systematic review," *Journal of the American Medical Directors Association*, vol. 18, no. 8, pp. 664–670, 2017.

[13] J. Ma, S. M. Lo, and W. G. Song, "Cellular automaton modeling approach for optimum ultra high-rise building evacuation design," *Fire Safety Journal*, vol. 54, pp. 57–66, 2012.

[14] M. Manley, Y. S. Kim, K. Christensen, and A. Chen, "Modeling emergency evacuation of individuals with disabilities in a densely populated airport," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2206, no. 1, pp. 32–38, 2011.

[15] K. Butler, E. Kuligowski, S. Furman, and R. Peacock, "Perspectives of occupants with mobility impairments on evacuation methods for use during fire emergencies," *Fire Safety Journal*, vol. 91, pp. 955–963, 2017.

[16] J. Koo, Y. S. Kim, B.-I. Kim, and K. M. Christensen, "A comparative study of evacuation strategies for people with disabilities in high-rise building evacuation," *Expert Systems with Applications*, vol. 40, no. 2, pp. 408–417, 2013.

[17] A. Sagun, D. Bouchlaghem, and C. J. Anumba, "Computer simulations vs. building guidance to enhance evacuation performance of buildings during emergency events," *Simulation Modelling Practice and Theory*, vol. 19, no. 3, pp. 1007–1019, 2011.

[18] E. Ronchi and D. Nilsson, "Modelling total evacuation strategies for high-rise buildings," *Building Simulation*, vol. 7, no. 1, pp. 73–87, 2014.

[19] A. Soltanzadeh, M. Alaghmandan, and H. Soltanzadeh, "Performance evaluation of refuge floors in combination with egress components in high-rise buildings," *Journal of Building Engineering*, vol. 19, pp. 519–529, 2018.

[20] Y. Ding, L. Yang, F. Weng, Z. Fu, and P. Rao, "Investigation of combined stairs elevators evacuation strategies for high rise buildings based on simulation," *Simulation Modelling Practice and Theory*, vol. 53, pp. 60–73, 2015.

[21] T. Fang, J. Yu, and J. Wang, "Study of staircase design effects on evacuation in architectural plane design," *Journal of Applied Fire Science*, vol. 22, no. 1, pp. 69–80, 2012.

[22] Y. Ding and L. Yang, "Occupant evacuation process study of public buildings based on computer modeling and simulation," *Journal of Applied Fire Science*, vol. 23, no. 3, pp. 365–380, 2013.

[23] Y. Q. Bao, "Study on fire prevention performance-based design of a large underground Banquet Hall," *Applied Mechanics and Materials*, vol. 94–96, pp. 2065–2069, 2011.

[24] C. Boje and H. Li, "Crowd simulation-based knowledge mining supporting building evacuation design," *Advanced Engineering Informatics*, vol. 37, pp. 103–118, 2018.

[25] L. Tan, M. Hu, and H. Lin, "Agent-based simulation of building evacuation: combining human behavior with predictable spatial accessibility in a fire emergency," *Information Sciences*, vol. 295, pp. 53–66, 2015.

[26] Y. Han, H. Liu, and P. Moore, "Extended route choice model based on available evacuation route set and its application in crowd evacuation simulation," *Simulation Modelling Practice and Theory*, vol. 75, pp. 1–16, 2017.

[27] G. Hofinger, R. Zinke, and L. Künzer, "Human factors in evacuation simulation, planning, and guidance," *Transportation Research Procedia*, vol. 2, pp. 603–611, 2014.

[28] Y. Li, H. Liu, G. -P. Liu, L. Li, P. Moore, and B. Hu, "A grouping method based on grid density and relationship for crowd evacuation simulation," *Physica A: Statistical Mechanics and Its Applications*, vol. 473, pp. 319–336, 2017.

[29] M. Kobes, I. Helsloot, B. de Vries, and J. G. Post, "Building safety and human behaviour in fire: a literature review," *Fire Safety Journal*, vol. 45, no. 1, pp. 1–11, 2010.

[30] M. F. Sherman, M. Peyrot, L. A. Magda, and R. R. M. Gershon, "Modeling pre-evacuation delay by evacuees in world trade center towers 1 and 2 on september 11, 2001: a revisit using regression analysis," *Fire Safety Journal*, vol. 46, no. 7, pp. 414–424, 2011.

[31] Y. Ma, L. Li, H. Zhang, and T. Chen, "Experimental study on small group behavior and crowd dynamics in a tall office building evacuation," *Physica A: Statistical Mechanics and Its Applications*, vol. 473, pp. 488–500, 2017.

[32] D. Li and B. Han, "Behavioral effect on pedestrian evacuation simulation using cellular automata," *Safety Science*, vol. 80, pp. 41–55, 2015.

[33] M. Liu and S. M. Lo, "The quantitative investigation on people's pre-evacuation behavior under fire," *Automation in Construction*, vol. 20, no. 5, pp. 620–628, 2011.

[34] S.-H. Wang, W.-C. Wang, K.-C. Wang, and S.-Y. Shih, "Applying building information modeling to support fire safety management," *Automation in Construction*, vol. 59, pp. 158–167, 2015.

[35] J. Choi, J. Choi, and I. Kim, "Development of BIM-based evacuation regulation checking system for high-rise and complex buildings," *Automation in Construction*, vol. 46, pp. 38–49, 2014.

[36] J. Ma, J. Chen, Y.-J. Liao, and L. Siuming, "Efficiency analysis of elevator aided building evacuation using network model," *Procedia Engineering*, vol. 52, pp. 259–266, 2013.

[37] H. Kim and S. Han, "Crowd evacuation simulation using active route choice model based on human characteristics," *Simulation Modelling Practice and Theory*, vol. 87, pp. 369–378, 2018.

[38] L. Zhang, Y. Wang, H. Shi, and L. Zhang, "Modeling and analyzing 3D complex building interiors for effective evacuation simulations," *Fire Safety Journal*, vol. 53, pp. 1–12, 2012.

[39] Y. Wu, J. Kang, and C. Wang, "A crowd route choice evacuation model in large indoor building spaces," *Frontiers of Architectural Research*, vol. 7, no. 2, pp. 135–150, 2018.

[40] J. Kang, I.-J. Jeong, and J.-B. Kwun, "Optimal facility-final exit assignment algorithm for building complex evacuation," *Computers & Industrial Engineering*, vol. 85, pp. 169–176, 2015.

[41] L. Chen, T.-Q. Tang, H.-J. Huang, J.-J. Wu, and Z. Song, "Modeling pedestrian flow accounting for collision avoidance during evacuation," *Simulation Modelling Practice and Theory*, vol. 82, pp. 1–11, 2018.

[42] A. U. K. Wagoum and A. Seyfried, "Conception, development, installation and evaluation of a real time evacuation assistant for complex buildings," *Procedia—Social and Behavioral Sciences*, vol. 104, pp. 728–736, 2013.

[43] J. Koo, B.-I. Kim, and Y. S. Kim, "Estimating the effects of mental disorientation and physical fatigue in a semi-panic evacuation," *Expert Systems with Applications*, vol. 41, no. 5, pp. 2379–2390, 2014.

[44] E. Eggert and F. Huss, "Medical and biological factors affecting mortality in elderly residential fire victims: a narrative review of the literature," *Scars, Burns & Healing*, vol. 3, Article ID 960007024, 2017.

[45] E. Kuligowski, R. Peacock, E. Wiess, and B. Hoskins, "Stair evacuation of older adults and people with mobility impairments," *Fire Safety Journal*, vol. 62, pp. 230–237, 2013.

[46] D.-j. Noh, J. Koo, and B.-I. Kim, "An efficient partially dedicated strategy for evacuation of a heterogeneous population," *Simulation Modelling Practice and Theory*, vol. 62, pp. 157–165, 2016.

[47] J. Bendel and H. Klüpfel, *Accessibility and Evacuation Planning—Similarities and Differences. Pedestrian and Evacuation Dynamics*, Springer Science+Business Media, Boston, MA, USA, 2011.

[48] K. Christensen and Y. Sasaki, "Agent-based emergency evacuation simulation with individuals with disabilities in the population," *Journal of Artificial Societies and Social Simulation*, vol. 11, no. 3, p. 9, 2008.

[49] D. T. Butry, R. E. Chapman, A. L. Huang, and D. S. Thomas, "A life-cycle cost comparison of exit stairs and occupant evacuation elevators in tall buildings," *Fire Technology*, vol. 48, no. 2, pp. 155–172, 2012.

[50] Y. J. Liao, S. M. Lo, J. Ma, S. B. Liu, and G. X. Liao, "A study on people's attitude to the use of elevators for fire escape," *Fire Technology*, vol. 50, no. 2, pp. 363–378, 2014.

[51] Y. Chen, J. Chen, Z. Fu, S.-D. Jiang, and L. Chen, "Gas characteristics and effectiveness of smoke control systems in elevator lobbies during elevator evacuation in a high-rise building fire," *Combustion Science and Technology*, vol. 190, no. 7, pp. 1232–1245, 2018.

[52] Y.-J. Liao, G.-X. Liao, and S.-M. Lo, "Influencing factor analysis of ultra-tall building elevator evacuation," *Procedia Engineering*, vol. 71, pp. 583–590, 2014.

[53] J. Chen, J. Ma, and S. M. Lo, "Event-driven modeling of elevator assisted evacuation in ultra high-rise buildings," *Simulation Modelling Practice and Theory*, vol. 74, pp. 99–116, 2017.

[54] E. Heyes and M. Spearpoint, "Lifts for evacuation-human behaviour considerations," *Fire and Materials*, vol. 36, no. 4, pp. 297–308, 2012.

[55] K. Andrée, D. Nilsson, and J. Eriksson, "Evacuation experiments in a virtual reality high-rise building: exit choice and waiting time for evacuation elevators," *Fire and Materials*, vol. 40, no. 4, pp. 554–567, 2016.

[56] N. Ding, T. Chen, and H. Zhang, "Simulation of high-rise building evacuation considering fatigue factor based on cellular automata: a case study in China," *Building Simulation*, vol. 10, no. 3, pp. 407–418, 2017.

[57] N. Ding, H. Zhang, and T. Chen, "Simulation-based optimization of emergency evacuation strategy in ultra-high-rise buildings," *Natural Hazards*, vol. 89, no. 3, pp. 1167–1184, 2017.

[58] T. Sano, E. Ronchi, Y. Minegishi, and D. Nilsson, "A pedestrian merging flow model for stair evacuation," *Fire Safety Journal*, vol. 89, pp. 77–89, 2017.

[59] W. Li, Y. Li, P. Yu et al., "Modeling, simulation and analysis of the evacuation process on stairs in a multi-floor classroom building of a primary school," *Physica A: Statistical Mechanics and Its Applications*, vol. 469, pp. 157–172, 2017.

[60] H. Qiu and S. Xiong, "New Hick's law based reaction test app reveals "information processing speed" better identifies high falls risk older people than "simple reaction time"" *International Journal of Industrial Ergonomics*, vol. 58, pp. 25–32, 2017.

[61] H. Liu, B. Xu, D. Lu, and G. Zhang, "A path planning approach for crowd evacuation in buildings based on improved artificial bee colony algorithm," *Applied Soft Computing*, vol. 68, pp. 360–376, 2018.

Hindawi
Discrete Dynamics in Nature and Society
Volume 2020, Article ID 4142824, 10 pages
https://doi.org/10.1155/2020/4142824

Research Article

Research on Partial Least Squares Method Based on Deep Confidence Network in Traditional Chinese Medicine

Wang-ping Xiong,[1] Tian-ci Li,[1] Qing-xia Zeng,[1] Jian-qiang Du,[1] Bin Nie,[1] Chih-Cheng Chen ,[2] and Xian Zhou[1]

[1]*School of Computer, Jiang Xi University of Traditional Chinese Medicine, Nanchang, Jiangxi, China*
[2]*School of Information and Engineering College, Jimei University, Xiamen, Fujian 361021, China*

Correspondence should be addressed to Chih-Cheng Chen; 3343033397@qq.com

Received 21 February 2020; Accepted 4 May 2020; Published 6 June 2020

Guest Editor: Fuqiang Gu

Partial least squares method has many advantages in multivariate linear regression modeling, but its internal cross-checking method will lead to a sharp reduction of the principal component, thereby reducing the accuracy of the regression equation, and the selection of principal components about the traditional Chinese medicine data is particularly sensitive. This paper proposes a kind of partial least squares method based on deep belief nets. This method mainly uses the deep learning model to extract the upper-level features of the original data, putting the extracted features into the partial least squares model for multiple linear regression and evading the problem that selects the number of principal components, continuously adjusting the model parameters until satisfied well-pleased accuracy condition. Using Dachengqitang experimental data and data sets in the UCI Machine Learning Repository, the experimental results show that the partial least squares analysis method based on deep belief nets has good adaptability to TCM data.

1. Introduction

The mechanism of treating diseases with traditional Chinese medicine is that multiple components of TCM correspond to multiple targets of diseases, so there are many related variables in TCM data. In the dose-effect relationship analysis of TCM prescriptions, the content of TCM prescriptions should be used to analyze the corresponding curative effects. To analyze the effects of independent variables on dependent variables, how to accurately select the number of independent variables is the key to establish a good dose-effect relationship model to achieve the goal of the better dose-effect analysis.

Partial least squares (PLS) [1, 2] is a multivariate statistical data analysis method, which integrates the basic functions of principal component analysis(PCA), canonical correlation analysis (CCA), and multiple linear regression (MLR) analysis. PLS simplifies the data structure by principal component extraction, selects the first several principal components by truncation, and only uses these principle components to obtain a model with better prediction performance. If the subsequent principal components can no longer provide more meaningful information for the interpretation of the model, adopting too many principal components will instead reduce the precision of regression fitting. Therefore, the selection of the number of principal components is particularly important. However, there are many related variables in TCM data. The cross-validation method will drastically reduce the number of principal components and reduce the accuracy of the regression equation. Deep belief nets (DBN) [3] are a multilayer neural network that fit the input data as much as possible. They extract the upper-level features of the original data through deep learning of the existing model, avoid the problem of selecting the number of principal components, convert the original m-dimensional input data into n-dimensional output data through the deep network, and the output data are another deep nonlinear expression of the original input

data. In 2011, Zhou [4] proposed embedding the fuzzy neural network model into the iterative form of partial least squares, which avoids the problem of principal component selection while achieving the nonlinear mapping effect, but the model results are easily affected by membership functions. In 2013, Qin [5] proposed a kernel partial least squares method, which combines intrinsic dimension estimation with the kernel partial least squares method and maps data to the high-dimensional linear space by the kernel function. Although the algorithm can avoid the selection of principal components and the nonlinear structure contained in the reaction sample data, the selection of kernel function is extremely difficult. In 2017, Zhu and JianqiangDu [6] proposed a partial least squares method integrated with the restricted Boltzmann machine (RBM-PLS), which extracts the required upper-layer features through the neural network structure, avoiding the problem of principal component selection, but the model results are affected by the initial values.

Therefore, this paper integrates the deep belief nets into the partial least squares, extracts the upper-level features of the original data through the deep learning model, avoids the problem of selecting the number of principal components, thus carries out multiple linear regression, and finally restores to the regression equation of the original variables. Through the verification of the test data, the parameters of the deep learning model and the number of hidden layers are adjusted in turn, finally achieving the purpose of improving the accuracy of the regression equation. The algorithm not only retains PLS's characteristic of eliminating multiple linear correlations of traditional Chinese medicine data but also can establish a good regression model for data with a small sample size. While avoiding the problem of selecting the number of principal components, it makes up for the defect that traditional Chinese medicine data are particularly sensitive to the selection of principal components, thus establishing a regression model suitable for traditional Chinese medicine characteristics.

2. Partial Least Squares

Partial least squares [7] is a new multivariate statistical data analysis method. Different from traditional multivariate regression statistics, it mainly studies the regression modeling of multiple dependent variables to multiple independent variables [8]. Regression modeling can be carried out when the number of sample points is less than the number of variables or multiple correlations exist between variables, and the regression coefficient of each variable is easy to explain.

The basic idea of partial least squares regression is that there is a set of independent variables $X = (x_1, x_2, x_3, L\ \ x_n)$ and dependent variables $Y = (y_1, y_2, y_3, L\ \ y_n)$, and the principal components (t_1 and u_1) extracted from independent variables and dependent variables are required to carry the variation information of their respective data tables to the maximum extent-var$(t_1) \longrightarrow$ max, var$(u_1) \longrightarrow$ max, they are required to have the greatest correlation degree $r(t_1, u_1) \longrightarrow$ max, multiple linear regression is then performed, and if the regression equation reaches satisfactory accuracy, the algorithm is terminated; otherwise, the second round of principal component extraction is continued from the residual information, thus reciprocating until satisfactory accuracy is reached.

3. Deep Belief Nets

Hinton and Salakhutdinov [9] proposed that restricted Boltzmann machines (RBMs) can be stacked greedily for training in 2016, thus forming deep belief nets (DBN). DBN are a probability generation model, which make the whole neural network generate training data according to the maximum probability through the weights between neurons in the training network structure, thus forming high-level abstract features [10].

DBN consist of two parts of neurons, namely, visible layer neurons and hidden layer neurons. The key component is RBM; feature extraction of input data can be realized by combining multiple layers of RBM. RBM is composed of two layers of neurons, visible layer and hidden layer, respectively. The specific structure diagram of RBM is shown in Figure 1.

Assuming that the RBM is n-layer, a DBN model that can extract features abstractly is constructed. v and h represent visible layer neurons and hidden layer neurons, respectively. The specific construction process is as follows: firstly, the weight and offset of the first trained RBM are fixed, the state of its hidden layer is taken as the input of the second RBM, the second RBM is trained and stacked on the first RBM, and so on to repeat the above process. The final model is shown in Figure 2.

Since DBN are formed by stacking multiple RBM, the main process of training DBN is the process of training RBM.

The number of neurons in the visible layer is defined as m, the number of neurons in the hidden layer is n, the state vector of the visible layer neurons is V, and the state vector of the hidden layer neurons is H. For a given set of state vectors, the energy of an RBM can be expressed by the following function [11]:

$$E(v, h \mid \theta) = -\sum_{i-1}^{n} a_i v_i - \sum_{j-1}^{m} b_j h_j - \sum_{i-1}^{n}\sum_{j-1}^{m} b_j w_{ij} v_i h_j. \quad (1)$$

Among them, the parameter of the model is $\theta = \{a_i, b_i, W_{ij}\}$, a_i is the offset of the visible layer neurons i,b_i is the offset of the hidden layer neurons j, and W_{ij} is the weight between the visible layer neurons and the hidden layer neurons j.

In the process of RBM training, the values of the parameter $\theta$are obtained through continuous iteration (the maximum parameter is obtained by using the stochastic gradient descent method), and the end of the iteration is determined by the effect of the fitted training data.

For trained DBN, under the condition of known parameter $\theta = \{a_i, b_i, W_{ij}\}$, preprocess the original data to obtain $V = (v_1, v_2, \ldots, v_n)$ and then put the data into the model, namely,

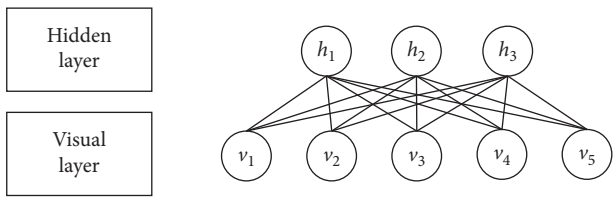


Figure 1: RBM structure diagram.

$$H = \text{sigmoid}(v \times w + a). \tag{2}$$

The original data are mapped into hidden layer data $H = (h_1, h_2, ..., h_m)$ through sigmoid function. In this way, the original data can be transformed into another deep expression, which achieves the purpose of extracting upper-level features.

4. Partial Least Squares Construction Based on DBN

In partial least squares, principal component analysis (PCA) and canonical correlation analysis (CCA) are mainly used for principal component extraction. It is not necessary to construct all the components for regression modeling. Usually, we use the cross-validity method to determine the number of principal components. When $Q_h^2 \geq 0.0975$, the added component is effective, and the regression model can be improved; that is, when $\sqrt{\text{PRESS}_h} \leq 0.95\sqrt{\text{SS}_{h-1}}$, the added component is regarded as beneficial and otherwise invalid.

$$Q_h^2 = 1 - \frac{\sum_{j=1}^{q} \text{PRESS}_{hj}}{\sum_{j=1}^{q} \text{SS}_{(h-1)j}} = 1 - \frac{\text{PRESS}_h}{\text{SS}_{h-1}} \geq 0.0975, \tag{3}$$

where h represents the number of components, PRESS_h represents the sum of squared prediction errors of equation Y, and SS_h represents the sum of squared errors of Y fitted with h components. Obviously, the value of PRESS_h is greater than SS_h, but less than SS_{h-1}. In fact, the square roots of the three are not much different, making the condition of $\sqrt{\text{PRESS}_h} \leq 0.95\sqrt{\text{SS}_{h-1}}$ very easy to meet. This leads to a sharp decrease in the number of selected components, thereby reducing the accuracy of the regression equation.

The feature extraction of the DBN network is mainly divided into two aspects: pretraining and parameter tuning. First, pretrain the network. RBM is the basic structure of the DBN network, so training is performed layer by layer in RBM. The standardized data set is used as the input of the underlying RBM, and the first layer feature h_1 of the data set is obtained through the previously trained parameter θ. Taking feature h_1 as the input data of the next layer of RBM, we can get feature h_2 which is more abstract than feature h_1. Through iterative abstraction, the final feature h_n is obtained. At this time, h_n integrates the different features abstracted by the underlying RBM, making the features more representative of the variables in the original data.

At this time, no matter whether there is a strong correlation between the original variables, different features can be extracted through the DBN network. Furthermore, according to the different functions of variables, they can be combined into more abstract and upper-level features that can represent different types of variables. In the end, the contribution of each independent variable to the experiment of the traditional Chinese medicine is brought into full play. Therefore, the DBN-PLS method proposed in this paper is to use DBN to replace the main component extraction part of PLS. The upper-level features of the original data extracted by DBN are used as the principal components, then the PLS external model is used for regression modeling, and finally the original variable regression equation is restored. The verification of the test data, in turn, adjusts the parameters of the deep learning model and the number of hidden layers. The goal of improving the accuracy of the regression equation is finally achieved.

Let the independent variable set $X = (x_1, x_2, x_3, L\, x_p)$ and dependent variable set $Y = (y_1, y_2, y_3, L\, y_q)$; X are matrices of $n * p$, and Y are matrices of$n * q$.

The specific construction process is as follows:

(1) Data preprocessing: standardized pretreatment of X and Y, respectively, to obtain E_0 and F_0.

(2) Independent $E_0 = (E_1, E_2, E_3, ..., E_p)$ and dependent $F_0 = (F_1, F_2, F_3, ..., F_q)$ variables are, respectively, put into the DBN model for training, and the processing steps are as follows:

  (1) According to the number of characteristic attributes of the independent variable $E_0 = (E_1, E_2, E_3, ..., E_p)$, the number of visible layer neurons in the DBN is determined, p. Since the main purpose is to reduce the dimension of characteristics, the number of hidden layers is generally smaller than the number of neurons in the visual layer. Here, the number is $p_1\,(p_1 < p)$.

  (2) Random initialization parameter $\theta = \{a_i, b_j, w_{ij}\}$. The offset vectors of the visible layer and the hidden layer are $a = \{a_1, a_2, ..., a_p\}$ and $b = \{b_1, b_2, ..., b_{p1}\}$, respectively, and the weight matrix is $W = \{W_{ij} | 0 \leq i \leq p, 0 \leq j \leq p_1\}$.

  (3) Take independent variables $E = \{E_1, E_2, ..., E_p\}$ as the input of the visual layer, train each RBM layer by layer, the hidden layer of the previous RBM is the visual layer of the next RBM and the output of the prethe input of the next RBM. In the training process, it is necessary to fully train the RBM of the previous layer before training the RBM of the current layer until the last layer.

(3) Take the hidden layer of the last layer of DBN $T = \{t_1, t_2, ..., t_k\}$ and $U = \{u_1, u_2, ..., u_1\}$ (k and l are the number of principal components extracted for independent variables and dependent variables, respectively) which are taken as principal components in PLS.

(4) After calculating the principal components in step 3, we put the principal components into the partial least squares external model for multiple linear regression analysis. Finally, the coefficient is reduced to a multiple linear regression equation of Y with respect to X. According to two evaluation indexes of

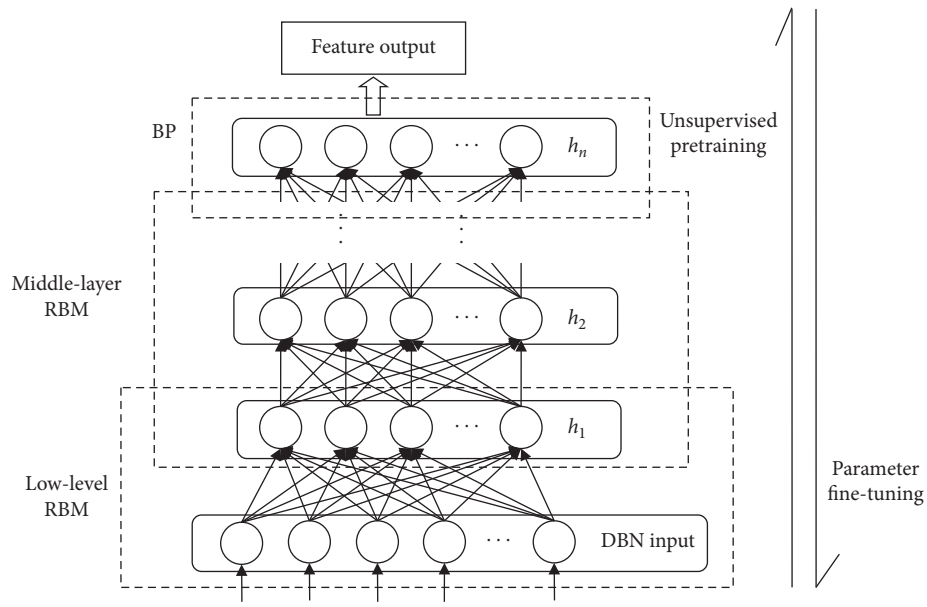


Figure 2: DBN structure diagram.

root mean square error (RMSE) and redetermination coefficient (R^2), it is judged whether the model is satisfied with the precision requirement at this time. If the requirement is met, the algorithm is stopped. If not, the superparameters of the model are continuously adjusted until the requirement is met.

As shown in Algorithm 1, the DBN-PLS model constructed in this paper is mainly composed of two parts: feature extraction and nonlinear PLS regression. Among them, DBN use the characteristics of their interlayer transmission to automatically extract features from the original data, and PLS is responsible for the regression of the extracted features. The model retains the characteristic that PLS can solve the problem of multiple correlations of traditional Chinese medicine data, while also avoiding the defect of cross-validity of PLS. It first performs canonical correlation analysis and multiple linear regression and then restores the regression equation of the original variables. Through the verification with the test data, the deep learning layer and parameter settings are adjusted in turn. Finally, the purpose of improving the accuracy of the final regression equation is achieved. The overall structural model is shown in Figure 3.

5. Experiment Results and Analysis

The experimental data in this paper are mainly from Dachengqitang experimental data (DCQT) of the Key Laboratory of Jiangxi University of Traditional Chinese Medicine and Housing, AirQuality, and CBM [12] on UCI data sets.

5.1. Experimental Data Description. The experimental data of Dachengqitang consist of 9 samples, which are the effects of active ingredients on pharmacological indexes in rat plasma at 9 different doses. Independent variables are the main active components in rat plasma, namely, emodin, rhein, chrysophanol, aloe-emodin, emodin methyl ether, magnolol, honokiol, hesperidin, and hesperetin; the dependent variables are the pharmacological indexes examined, namely, first defecation time, motilin, and vasoactive intestinal peptide. Some experimental data are shown in Table 1.

For UCI data, Housing, AirQuality, and CBM data sets are selected, with sample sizes of 506, 9,357, and 11,934, respectively. See http://archive.ics.uci.edu/ml/ for a detailed description.

5.2. Data Standardization. Data normalization (normalization) processing is a necessary task when to start data mining. Different data have different dimensions and dimension units, and this situation will affect the results of data analysis. To eliminate the dimensional impact between data and the data itself, data standardization is needed to resolve the comparability of data metrics. In order to comprehensively compare and evaluate, the above statistical data are separately standardized. After the original data are standardized, the indicators reach the same level, which is suitable for comparative evaluation [4]. The standardized method used in this paper is Z-score standardization, also called standard deviation standardization. The processed data conform to the standard normal distribution, that is, the mean is 0, and the variance is 1. Its conversion function is

$$x^* = \frac{x - \mu}{\sigma}, \tag{4}$$

where $\mu = (1/n)\sum_{i=1}^{n} x_i$ represents the mean of each feature of data X and $\sigma = \sqrt{(1/n-1)\sum_{i=1}^{n}(x_i - \mu)^2}$ represents the standard deviation of each feature of data X.

Taking Dachengqitang data as an example, the data set is subjected to data standardization processing. The results are shown in Table 2.

Input: original sample data set Dataset (D);
Output: DBN-PLS equation.
Step 1 preprocesses the data to obtain (E_0, F_0)
Step 2 deep belief nets (DBN)
Initialize model parameters $\theta = \{a_i, b_i, W_{ij}\}$
Layersize = 1
hidden_layers_sizes = [4, 4, 4]
While Whether the number of RBM layer size reaches the precision condition
while Whether the number *layersize* of each layer of neurons reaches the accuracy condition
for z in *layersize*
Calculating the probability that hidden layer neurons are activated $P_z(h_i|v_i)$
Take Gibbs sampling to extract a sample: $v_{i+1} \sim P_z(v_{i+1}|h_i)$
Reconstruct the visual layer with h_i is used to calculate the probability of activated hidden layer neurons $P_z(v_{i+1}|h_i)$
Take Gibbs sampling to extract a sample: $v_{i+1} \sim P_z(v_{i+1}|h_i)$
V_{i+1} is used to calculate the probability of activated hidden layer neurons $P_z(h_{i+1}|v_{i+1})$
Update weight
$$b_i \longleftarrow b_i + \lambda(v_i - v_{i+1})$$
Step 3 The characteristics extracted by independent variables and dependent variables are calculated by the DBN model, respectively. $T = sigmoid(E_o \times W^t + a^t)$, $U = sigmoid(F_o \times W^u + a^u)$, put the PLS outer model into multiple linear regression, and find the DBN-PLS equation.
Denormalized reduction of coefficients Y multiple linear regression equations for X
Step 4 End

Algorithm 1: Partial least squares algorithm based on DBN.

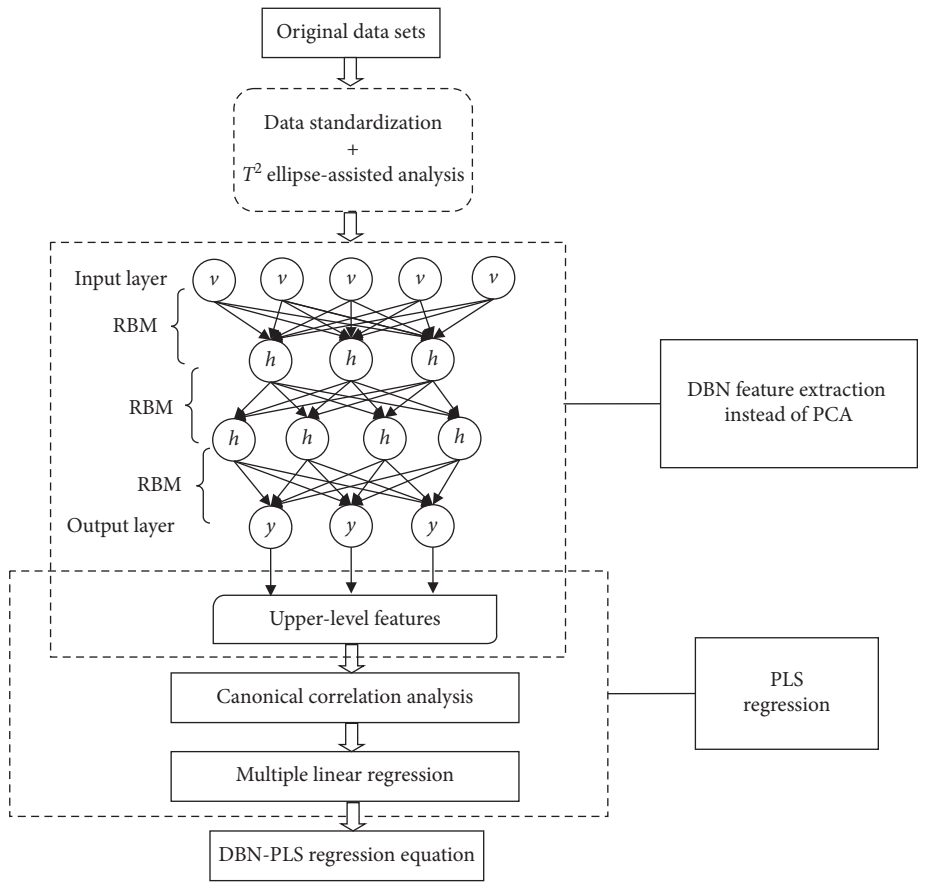


Figure 3: DBN-PLS model framework.

Table 1: Experimental data set of Dachengqitang (DCQT).

Emodin	Rhein	Chrysophanol	Aloe-emodin	Emodin methyl ether	Magnolol	Honokiol	Hesperidin	Hesperetin	First defecation time	Motilin	Vasoactive intestinal peptide
1.19	31.35	2.82	2	0.67	0.5	0.18	4.91	0.99	6.91	0.15	3.46
2.6	54	3.95	1.89	0.68	0.26	0.13	4.61	1.17	6.55	0.18	3.27
2.82	26.55	4.85	1.49	0.76	0.15	0.28	6.91	2.49	5.78	0.19	3.88
3.08	131.55	16.5	2.3	2.07	0.72	0.79	7.91	0.3	6	0.19	3.88
4.08	66.75	26	1.77	1.5	0.69	0.44	8.15	2.33	5.88	0.18	3.61
...	...	...	...	...	...	...	...	...	...	...	...

Table 2: Standardization processing results (DCQT).

Emodin	Rhein	Chrysophanol	Aloe-emodin	Emodin methyl ether	Magnolol	Honokiol	Hesperidin	Hesperetin	First defecation time	Motilin	Vasoactive intestinal peptide
-0.689	−0.741	−1.071	−0.479	−1.027	−0.603	−0.822	−1.153	−0.638	0.224	−1.492	−1.021
-0.563	−0.599	−1.014	−0.506	−1.022	−1.114	−0.876	−1.218	−0.550	−0.139	0.426	−1.335
-0.544	−0.771	−0.969	−0.606	−0.974	−1.348	−0.715	−0.721	0.093	−0.914	1.066	−0.327
-0.521	−0.115	−0.383	−0.403	−0.200	−0.135	−0.167	−0.504	−0.975	−0.693	1.066	−0.327
-0.432	−0.520	0.095	−0.536	−0.537	−0.199	−0.543	−0.452	0.015	−0.814	0.426	−0.773
...	...	...	...	...	...	...	...	...	...	...	...

5.3. Auxiliary Analysis. The T^2 ellipse-assisted analysis technique finds the singular points in the sample by analyzing the contribution rate of the sample points to the components. We define the contribution rate T_{hi}^2 of the i-th sample point to the h-th component t_h to be $T_{hi}^2 = t_{hi}^2/(n-1)s_h^2$, where t_{hi} is the projection value of the i-th sample point on the h-th main axis a_h and $s_h^2$is the variance of the component t_h. Then, the cumulative contribution rate $T_i^2$of the i-th sample point pair component $t_1, t_2, , t_m$ is $T_i^2 = \sum_{h=1}^{m} t_{hi}^2/(n-1)s_h^2$. When T_i^2 satisfies $T_i^2 \geq (m(n^2-1)/n^2(n-m))F_{0.05}(m, n-m)$, it is considered that, at the 95% test level, the cumulative contribution rate of the i-th sample point to the component $t_1, t_2, \ldots, t_m$ is too large, and the sample point i is called a singular point.

In particular, when $m = 2$,

$$T_i^2 = \left(\frac{t_{1i}^2}{(n-1)s_1^2} + \frac{t_{2i}^2}{(n-1)s_2^2}\right) \geq \frac{2(n^2-1)}{n^2(n-2)} F_{0.05}(2, n-2)$$

$$\frac{t_{1i}^2}{s_1^2} + \frac{t_{2i}^2}{s_2^2} \geq \frac{2(n-1)(n^2-1)}{n^2(n-2))} F_{0.05}(2, n-2). \tag{5}$$

Let

$$c = \frac{2(n-1)(n^2-1)}{n^2(n-2)} F_{0.05}(2, n-2), \tag{6}$$

that is,

$$\frac{t_{1i}^2}{s_1^2} + \frac{t_{2i}^2}{s_2^2} \geq c. \tag{7}$$

Equation (7) represents an area outside the ellipse $(t_{1i}^2/s_1^2) + (t_{2i}^2/s_2^2) = c$ range. In other words, the singular point is distributed outside the ellipse range; then, on the t_1/t_2-plane coordinate system, the T^2 ellipse and the scatter plot of the data set can be drawn. Those points that are outside the T^2 ellipse can be called singular points.

From the analysis of the four sets of data, from equation (7), when the extracted component m is 2, the right side of the inequality is an ellipse, and we can make an ellipse on the t_1/t_2 planar graph. Thus, the existence of singular points can be visually reflected in the graph. Therefore, in this paper, two principal components are extracted separately for the four models, and the singular points in the sample points in the model are discriminated according to discriminant conditional formula (7). Four T^2 ellipses were created using SIMCA-P software to show the presence of sample points, as shown in Figures 4–7.

From Figures 4–7, it can be concluded that there are no singular points in the DCQT data set, and there are 22, 544, and 645 singular points in the Housing, AirQuality, and CBM data sets, accounting for 4.3%, 5.8%, and 5.4% of the total samples, respectively. The sample points concentrated outside the ellipse take values away from the average level of the sample points, the so-called singular points. Find out these singular points and eliminate them, and then use the processed data to optimize the experimental analysis.

There are also some common methods in the singularity elimination phase of partial least squares. The direct decomposition algorithm (TDDA) [13] directly decomposes the multidimensional data set according to all its dimensions to obtain several one-dimensional data sets. Therefore, the multidimensional data are mapped to the one-dimensional space, and the purpose of dimensionality reduction is realized. The multidimensional scaling method (TMSM) [13] is a graphical method for visually representing a research object in a low-dimensional space based on the similarity between the research objects and performing clustering or dimensional inclusion analysis. The T^2 ellipsoid method [14] is a modification of the T^2 ellipse analysis technique, extending the two-dimensional space T^2 elliptical plan to the three-dimensional T^2 ellipsoid and then using the T^2 ellipsoid to

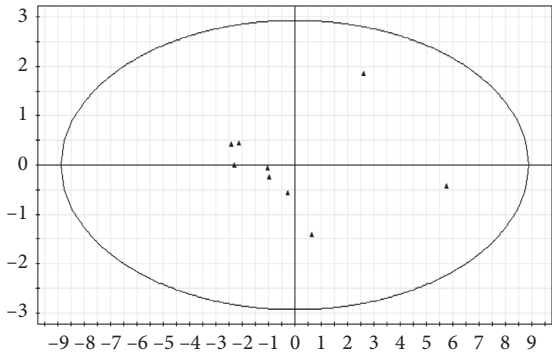


Figure 4: Singular point identification of DCQT.

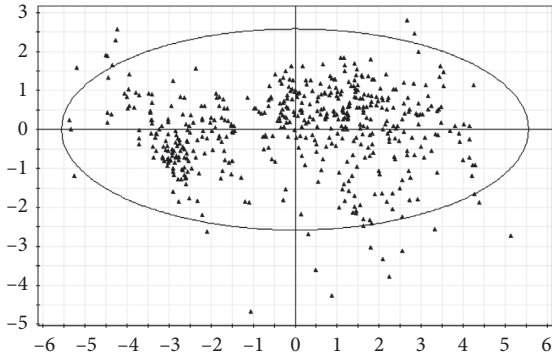


Figure 5: Singular point identification of Housing.

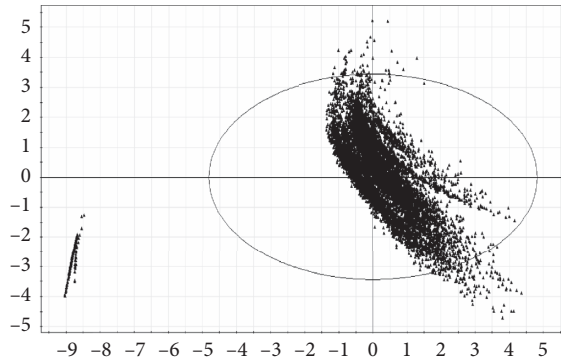


Figure 6: Singular point identification of AirQuality.

identify specific points. Taking the Housing data set as an example, the above method is used to identify the specific point of the data set. Figure 8 shows the graphical results of the T^2 ellipsoid method. The results of all methods are shown in Table 3.

Obviously, by analyzing the comparison results, it can be found that the Housing data set after T^2 ellipse cleaning has the highest fitting value of PLS. TDDA may cause omissions when detecting the specific points of multidimensional data, so some specific points will be ignored; when using TMSM, the effect of each dimension on the Euclidean distance is equal, but in practice, the fluctuation range of each dimension is not the same, and the TCM data are highly correlated, so it is not too suitable for this; the T^2 ellipsoid

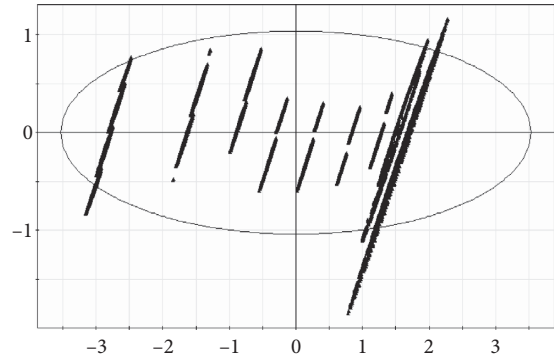


Figure 7: Singular point identification of CBM.

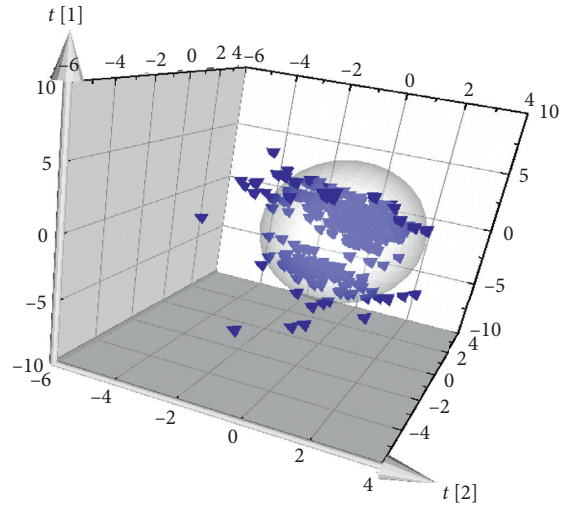


Figure 8: Results of the T^2 ellipsoid method for processing the Housing data set.

method detects some normal data as a specific point, which leads to a decrease of redetermination coefficient (R^2). From there, it can be judged that the T^2 ellipse analysis method has a better performance in the recognition of specific points.

5.4. Experimental Process and Result Analysis. In order to verify the feasibility and effectiveness of the DBN-PLS method, three deep belief net models are, respectively, set to train the number of hidden layers. The number of hidden layers in the first DBN model is set to 3. The number of hidden layers in the second DBN model is set to 4. The number of hidden layers in the third DBN model is set to 5. The number of neurons in the hidden layer of the above three DBN models depends on the specific situation. Each DBN model is combined with PLS, and the above three situations are analyzed and compared with the original partial least squares (PLS) and the partial least squares (RBM-PLS) integrated into the restricted Boltzmann machine in the literature [12]. Four groups of experimental data are used for comparison and verification. The specific description of the data set is shown in Table 4.

In the specific process of the experiment, the model is optimized by adjusting the model parameters, and the effects

Table 3: Results of the specific point identification method of the Housing data set.

Data name	Original number of records	Number of specific points	Cleaned data	R^2
T^2 ellipse method	0.9687	0.6934	0.2797	0.3479
TDDA	40.9841	41.2391	33.4516	40.1487
TMSM	74.5369	76.1351	17.5692	18.3622
T^2 ellipsoid method	0.0634	0.0165	0.0075	0.0076

Table 4: Data set description.

Data name	Number of samples	Number of independent variables	Number of dependent variables
DCQT	9	9	3
Housing	506	11	3
AirQuality	9357	10	3
CBM	11934	16	2

of several methods are compared at the same level of the learning and training set. Root mean square error (RMSE) and redetermination coefficient (R^2) were investigated, respectively. The experimental results are shown in Table 5.

According to the experimental results in Table 5, RMSE and R^2 of DBN-PLS are better than PLS and RBM-PLS. When the number of hidden layers (i.e., the number of RBM overlays) of the DBN-PLS method is 3, the effect is the best, the number of hidden layers increases continuously, and the effect also decreases gradually. The specific analysis process is as follows.

In the experimental data of Dachengqitang, the RMSE of PLS is much larger than that of RBM-PLS and DBN-PLS, and the RMSE of RBM-PLS is larger than that of DBN-PLS. When the number of hidden layers of the DBN-PLS method increases continuously, the RMSE increases continuously, that is, the effect decreases continuously. The RMSE of the five methods is 0.9687, 0.6934, 0.2797, 0.3420, and 0.3479, respectively. The R^2 value of DBN-PLS is greater than PLS and RBM-PLS, and the R^2 value of RBM-PLS is greater than PLS. When the number of hidden layers of the DBN-PLS method increases continuously, the R^2 value decreases continuously, that is, the effect decreases continuously. The R^2 values of the five methods are 0.6942, 0.7962, 0.9421, 0.9135, and 0.9105, respectively. When the number of RBM superposed by DBN-PLS is 3, 4, and 5, respectively, the values of RMSE and R^2 are 0.2797 and 0.9421, 0.3420 and 0.9135, and 0.3479 and 0.9105, respectively. Based on the above two evaluation indexes, it can be concluded that DBN-PLS is better than RBM-PLS and PLS for the experimental data of Dachengqitang, and RBM-PLS is better than PLS. When the number of RBM superposed by DBN-PLS is 3, the effect is most significant.

In the UCI standard data set, Housing is a medium data sample, while AirQuality and CBM are large data samples. Among all kinds of data, DBN-PLS has the best effect. In Housing and AirQuality data sets, RBM-PLS has the same effect as PLS, even slightly worse than PLS. For example, in the Housing data set, RMSE of RBM-PLS is larger than PLS, 41.2391 and 40.9841, respectively. Similarly, the R^2 value of RBM-PLS is less than PLS, 0.1813 and 0.1987, respectively. In the CBM data set, RMSE of RBM-PLS is less than PLS, 0.0165 and 0.0634, respectively. In contrast, the R^2 value of RBM-PLS is less than PLS, 0.5010 and 0.6935, respectively. The above effect is produced because the results of the RBM-PLS model are affected by the initial values. Different initial values have different effects, which will produce slightly worse effects than PLS. To sum up, DBN-PLS has the most significant effect on the above two evaluation indexes, whether it is the experimental data set of Dachengqi decoction or the UCI standard data set. Especially, when the number of RBM in the model is 3, the effect is the best. On the contrary, the superposition number increases, and the effect decreases. This shows that when DBN-PLS is used for feature extraction, the superposition number of RBM is 3, and the effect is the best.

To display the experimental results more intuitively, graphs are drawn to reflect the fluctuation of root mean square error (RMSE) and redetermination coefficient (R^2). Due to the different orders of magnitude of RMSE and R^2 of each data set, to compare the fluctuation of RMSE and R^2 of each data set in different methods conveniently, the experimental result data are centrally processed and mapped to [0,1], Figures 9 and 10 are drawn.

From Figures 9 and 10, it can be seen more intuitively that DBN-PLS has significantly improved effect in various indexes, and the effect is better than RBM-PLS and PLS. When the number of hidden layers of the DBN-PLS method is 3, the effect is the best. The number of hidden layers increases continuously, and the effect of each index shows a slow downward trend. In the DCQT data set, RBM-PLS has better performance than PLS, while in the UCI standard data set, RBM-PLS has slightly worse performance than PLS. This is because the results of the RBM-PLS model are affected by the initial values. The initial values selected are different, and the effects are different.

To sum up, among the above four groups of experiments, DBN-PLS has the best effect of multiple linear regression, which indicates that the fitting effect is better. As the number of hidden layers in DBN-PLS increases, the effect weakens, indicating that when DBN-PLS is used for feature extraction, the best effect is when the number of RBM overlays is 3. In different data sets, the effects of RBM-

Table 5: Comparison of experimental results.

Data name	PLS		RBM-PLS		DBN(3)-PLS		DBN(4)-PLS		DBN(5)-PLS	
	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
DCQT	0.9687	0.6942	0.6934	0.7962	0.2797	0.9421	0.3420	0.9135	0.3479	0.9105
Housing	40.9841	0.1987	41.2391	0.1813	33.4516	0.3680	37.2034	0.3017	40.1487	0.2330
AirQuality	74.5369	0.9247	76.1351	0.6209	17.5692	0.9487	18.0087	0.9460	18.3622	0.9439
CBM	0.0634	0.6935	0.0165	0.5010	0.0075	0.7940	0.0075	0.7919	0.0076	0.7879

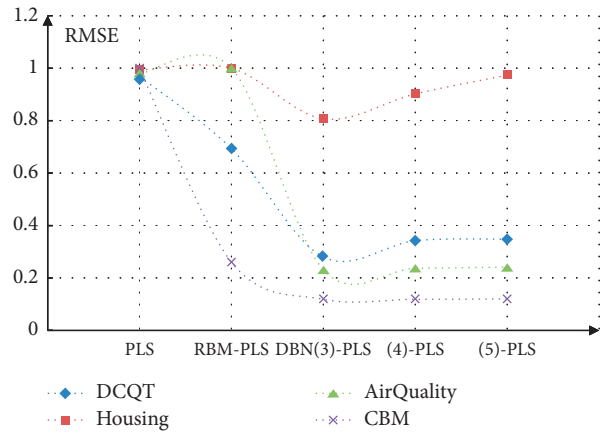


Figure 9: RMSE comparison of different methods under four groups of experimental data.

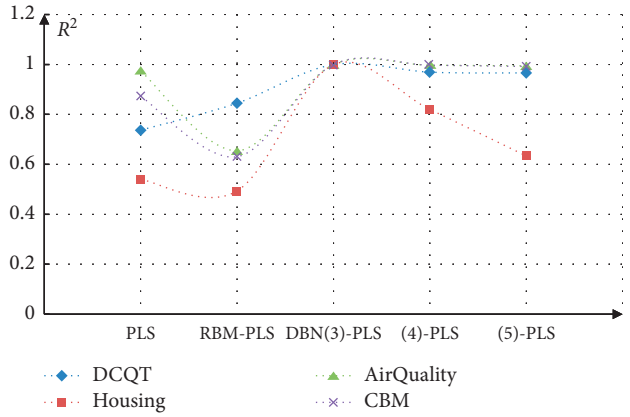


Figure 10: Comparison of methods R^2 under four groups of experimental data.

PLS and PLS are slightly different because the results of the RBM-PLS model are affected by initial values. To sum up, the partial least squares method based on DBN has good adaptability to TCM data.

6. Conclusions

The TCM data are modeled based on the partial least squares method. In addition to the same process as the modeling methods in other fields, there are also features such as large data volume, multiple correlations, nonlinearity, and random distribution of data. The main performance is ① Modeling samples selected for massive data affect the accuracy of the PLS model. ② Partial least squares method has limited effect in identifying and eliminating abnormal observation data records. ③ Limitations of principal component extraction methods.

Therefore, the work of this paper is improved based on the above partial least squares method:

(1) The experimental data of the traditional Chinese medicine: there are often many variables related to the dependent variable, and even the observation data of some independent variables are costly, but they can now be collected by leading-edge equipment. If such variables are included in the regression model, not only will the calculation amount be increased and the application cost of the model be increased but also the sample data are multidimensional, that is, a sample is characterized by multiple features. The dimensions and numerical values of these features are different, making the model unstable. Therefore, the work of this paper is firstly through the Z-score standardization method for data standardization processing compared with the min-max standardization method [15]; it is not necessary to redefine the values of max and min so that different features have the same scale. Thus, when learning parameters, different features have the same degree of influence on the parameters.

(2) High-quality data from drug trials are the basis for building analytical models. The experimental data are not only affected by the statistical distribution characteristics of the system parameters but also affected by the operating habits of the experimental personnel so that the number of occurrences of various dispensing groups is extremely uneven, and the observation data cannot contain large errors; otherwise, the result is not reliable. Therefore, the T^2 ellipse is used to observe the distribution of the sample points and the similarity structure on the t_1/t_2plan and to find those singular points whose values are far from the sample points combined with the average level. Achieve the purpose of eliminating abnormal data. Compared to TDDA, TMSM, and T^2 ellipsoid method mentioned by Sun Shuwei [16] and Cui Lizhen [17], the T^2 ellipse method can better eliminate noise and improve the value of the redetermination coefficient of PLS.

(3) In the partial least squares algorithm, if the number of principal components selected to participate in the regression is too small, some useful information in the original independent variable matrix is ignored,

which affects the fitting ability and prediction accuracy of the model; if the number of selected principal components is too large, although the fitting accuracy of the model can be improved, overfitting may occur due to introduction of irrelevant noise, resulting in a decrease in prediction accuracy. The most common method for extracting the number of components by partial least squares is to use the cross-validation (CV) [18] method to investigate the change of model prediction ability after adding new components, but the existing CV is insufficient. If the number of verification sets is too small, it is easy to overfit; the criteria for determining the optimal number of verification sets need to be given.

This paper proposes a partial least squares method based on DBN, which makes full use of the joint probability density distribution in the DBN generation model to learn the upper-layer features of the data and avoids the selection of the number of principal components. Combined with the characteristics of PLS, in the case of small sample size, it is still possible to regression modeling and maximize the relationship between independent variables and dependent variables and give full play to the advantages of the algorithms themselves. Experiments on Chinese medicine data and UCI standard data sets show that DBN-PLS significantly improves the accuracy of the regression equation and the expression of nonlinear structures. Compared with the conventional partial least squares method, it is more in line with the theoretical value and better solves the problem of determining the number of principal components in the partial least squares method.

Data Availability

The data in this article can be shared and used free of charge, but the data not listed in this article come from the Key Laboratory of traditional Chinese medicine preparation of Jiangxi University of Traditional Chinese Medicine, which involve patient privacy and business secrets and cannot be shared and used.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (nos. 61762051 and 61562045), the Jiangxi Province Science and Technology Plan Projects (nos. 20171BBG70108 and 20171ACE50021), and the Project of Jiangxi Health Planning Commission (nos. 20165537, 2017A284, 2017A313, and 20185516).

References

[1] Z. G. Shang, Y. H. Dong, and M. M. Li, "Robust feature selection and classification algorithm based on partial least square regression," *Computer Application Research*, vol. 37, no. 3, pp. 871–875, 2017.

[2] H. Abdi and L. Williams, B. Reisfeld and A. N. Mayeno, *Partial Least Squares Methods: Partial Least Squares Correlation and Partial Least Square Regression*, vol. 930, pp. 549–579, Humana Press, Totowa, NJ, USA, 2013.

[3] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[4] L. Zhou, *Research on Feature Extraction Based on Nonlinear Partial Least Square*, Nanjing University of Technology, Nanjing, China, 2011.

[5] J. C. Qin, "An optimal band selection method for hyperspectral images based on kernel partial least square method," *Journal of Photogrammetry*, vol. 30, no. 2, pp. 171–176, 2013.

[6] Z. P. Zhu and J. Q. JianqiangDu, "Partial least square optimization method integrating restricted Boltzmann machine," *Calculating Machine Works*, vol. 7, pp. 193–202, 2017.

[7] J. Magnanensi, F. Bertrand, M. Maumy-Bertrand, and N. Meyer, "A new universal resample-stable bootstrap-based stopping criterion for PLS component construction," *Statistics and Computing*, vol. 27, no. 3, pp. 757–774, 2017.

[8] G. A. F. Seber and A. J. Lee, *Linear Regression Analysis*, Wi-ley, Hoboken, NJ, USA, 2012.

[9] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[10] F. Y. Liu et al., "Review of deep confidence network models and applications," *Computer Engineering and Applications*, vol. 54, no. 1, pp. 11–19, 2018.

[11] X. S. Yang, *Nature Inspired Meta-Heuristic Algori-Thms*, Luniver Press, Frome, UK, 2010.

[12] UCI Machine Learning Repository, http://archive.ics.uci.edu/ml/datasets/Air+Qu.

[13] G. R. Zhang, *On the Detection Method Multidimensional Outliers*, Shenyang Sport University, Shenyang, China, 2011.

[14] J. X. Guo, *Study on Improved High-Dimension and Nonlinear Partial Least-Squares Regression Method and Applications*, Tianjin University, Tianjin, China, 2010.

[15] X. J. Ji, B. GuiS, and D. WangG, "Using min-max normalization to measure the differences of regional economic growth—a case study of yulin area, shanxi province," *Economy And Management*, vol. 30, no. 3, pp. 54–56, 2016.

[16] S. W. Sun, *Reconstruction Model of Support Vector Machine and Research on SMO Algorrthm*, Xidian University of Electronic Technology, Xi'an, China, 2009.

[17] L. Z. Cui, "Multidimensional scaling positioning algorithm based on RSSI in underground coal mine," *Safety in Coal Mines*, vol. 48, no. 6, pp. 108–111, 2017.

[18] C. Y. Sun and M. Y. Chang, "The study of beijing in water demand forecast model," *Journal of Applied Statistics and Management*, vol. 36, no. 6, pp. 1049–1058, 2017.

Hindawi
Discrete Dynamics in Nature and Society
Volume 2020, Article ID 4716286, 10 pages
https://doi.org/10.1155/2020/4716286

Research Article

Optimizing Biomedical Ontology Alignment through a Compact Multiobjective Particle Swarm Optimization Algorithm Driven by Knee Solution

Xingsi Xue,[1,2,3,4] **Xiaojing Wu**,[5] **and Junfeng Chen**[6]

[1]*Fujian Provincial Key Laboratory of Big Data Mining and Applications, Fujian University of Technology, Fuzhou 350118, China*
[2]*Intelligent Information Processing Research Center, Fujian University of Technology, Fuzhou 350118, China*
[3]*Fujian Key Lab for Automotive Electronics and Electric Drive, Fujian University of Technology, Fuzhou 350118, China*
[4]*Guangxi Key Laboratory of Automatic Detecting Technology and Instruments, Guilin University of Electronic Technology, Guilin, Guangxi 541004, China*
[5]*College of Information Science and Engineering, Fujian University of Technology, Fuzhou 350118, China*
[6]*College of IOT Engineering, Hohai University, Changzhou 213022, China*

Correspondence should be addressed to Xingsi Xue; jack8375@gmail.com

Received 19 January 2020; Revised 15 March 2020; Accepted 19 March 2020; Published 1 May 2020

Guest Editor: Chi-Hua Chen

Nowadays, most real-world decision problems consist of two or more incommensurable or conflicting objectives to be optimized simultaneously, so-called multiobjective optimization problems (MOPs). Usually, a decision maker (DM) prefers only a single optimum solution in the Pareto front (PF), and the PF's knee solution is logically the one if there are no user-specific or problem-specific preferences. In this context, the biomedical ontology matching problem in the Semantic Web (SW) domain is investigated, which can be of help to integrate the biomedical knowledge and facilitate the translational discoveries. Since biomedical ontologies often own large-scale concepts with rich semantic meanings, it is difficult to find a perfect alignment that could meet all DM's requirements, and usually, the matching process needs to trade-off two conflict objectives, i.e., the alignment's recall and precision. To this end, in this work, the biomedical ontology matching problem is first defined as a MOP, and then a compact multiobjective particle swarm optimization algorithm driven by knee solution (CMPSO-K) is proposed to address it. In particular, a compact evolutionary mechanism is proposed to efficiently optimize the alignment's quality, and a max-min approach is used to determine the PF's knee solution. In the experiment, three biomedical tracks provided by Ontology Alignment Evaluation Initiative (OAEI) are used to test CMPSO-K's performance. The comparisons with OAEI's participants and PSO-based matching technique show that CMPSO-K is both effective and efficient.

1. Introduction

Decision-making requires finding an optimum solution to a decision problem in the process of identifying and evaluating alternatives. Nowadays, most real-world decision problems consist of two or more incommensurable objectives to be optimized simultaneously, so-called multiobjective optimization problems (MOPs) [1]. In general, MOP does not have one optimum solution but a set of solutions, so-called Pareto optimal solutions, which are superior to the others in terms of one or more objectives. In many situations, a decision maker (DM) only prefers one single optimum solution in the Pareto front (PF) [2], and therefore, the optimization and the decision process are often combined. One of the multiobjective methods is driven by the PF's knee solution which is the one for which an improvement in one of its objectives will result in a deterioration in another. Often, PF's knee solution is logically preferred to the DM if there are no user-specific or problem-specific preferences [3].

In this context, the biomedical ontology matching problem in the Semantic Web (SW) domain is investigated. Although biomedical ontology is extensively used in the biomedical domain to model the biomedical knowledge, it is

developed and maintained by different communities, and the same biomedical knowledge could be described with different terminologies or in different contexts. To bridge the semantic gaps between two biomedical ontologies and support their communications, it is necessary to find their identical concept mapping, which is the so-called biomedical ontology matching. Since biomedical ontologies often own large-scale concepts with rich semantic meanings, it is difficult to find a perfect alignment that could meet all DM's requirements, and usually, the matching process needs to trade-off two conflict objectives, i.e., the alignment's recall and precision. To this end, in this work, the biomedical ontology matching problem is defined as a MOP. Although there are many studies on the single-objective approaches for addressing the ontology matching problem [4–8], the research on multiobjective ontology matching techniques is still in this infancy [9]. Being inspired by the success of the particle swarm optimization algorithm (PSO) in the ontology matching domain [10], large-scale discrete optimization domain [11–13], and biomedical engineering [14], in this work, a compact multiobjective particle swarm optimization algorithm driven by knee solution (CMPSO-K) is further proposed to address this problem. In particular, CMPSO-K uses a compact multiobjective evolutionary mechanism to efficiently optimize the alignment's quality and a max-min approach to determine the PF's knee solution. The contributions made in this paper are as follows: (1) a discrete multiobjective optimal model for the biomedical ontology matching problem is constructed; (2) a hybrid biomedical concept similarity metric is proposed, which can effectively calculate the similarity value of two biomedical ontology concepts; and (3) a CMPSO-K is proposed to optimize the biomedical ontology alignment, which takes into consideration both DM's preference and algorithm's performance.

The rest of the paper is organized as follows: Section 2 overviews the SIA-based ontology matching techniques; Section 3 defines the biomedical ontology matching problem and presents the similarity measure on biomedical concepts; Section 4 presents the CMPSO-K-based biomedical ontology matching technique in detail; Section 5 shows the experimental results; and finally, Section 6 draws the conclusion and presents the future work.

2. Swarm Intelligence Algorithm-Based Ontology Matching Technique

The first generation of swarm intelligence algorithm- (SIA-) based matcher dedicates to optimize the ways of aggregating various ontology matchers' corresponding alignments. The very first matching system is Genetics for Ontology ALignments (GOAL) [7], which uses evolutionary algorithm (EA) to optimize the aggregating weight set of different ontology matchers. Later, Alexandru-Lucian and Iftene [15] further used EA to optimize one more parameter to filter the unauthentic concept mappings to final alignment. Acampora et al. [4] introduced a local search process into EA's evolving process to improve the algorithm's performance. Xue and Wang [16] used a new metric as the fitness function to guide the algorithm's search direction. Their approach can address the holistic matching problem and determine a universal weight configuration for matching several pairs of ontologies at a time. He et al. [6] proposed to utilize artificial bee colony algorithm (ABC) to optimize all the parameters in the matching process, whose results are better than the EA-based matchers. More recently, Xue et al. [17] proposed a new approach that uses NSGA-III [18] to combine various similarity measures without tuning the aggregating parameters. However, when the scale of the similarity measures becomes huge, e.g., more than 50 similarity measures, this approach could be inefficient.

The above matchers need to calculate all the matchers' alignments and store them in the main memory before aggregating them, which requires huge memory consumption. Recently, the second generation of the SIA-based matcher tries to directly find an entity correspondence set that is close to the golden alignment. GAOM (genetic algorithm-based ontology matching) [8] regards two ontologies as two discrete concept sets and employs EA to determine the optimal entity mapping set. Alves et al. utilized a memetic algorithm (MA), which combines EA with a local search strategy, to execute the instance-based ontology matching process [5]. They first matched the instances and then propagated the instance pair's similarity value to the corresponding concepts. More recently, MapPSO [10] uses PSO to determine the optimal entity correspondence set. In particular, MapPSO introduces a new quality measure on the ontology alignment, which depends on the statistical results on the alignment. More recently, Chu et al. [19] first modeled the ontologies in the vector space so that two entities' similarity value can be calculated through the cosine function, and then EA was used to determine the optimal alignments. Xue [20] proposed a new similarity metric for measuring the biomedical concepts' similarity value and then used the firefly algorithm (FA) to optimize the biomedical ontology alignment. However, their proposal suffers from the premature convergence when matching large-scale ontologies. Our proposal also belongs to this category, but different from the existing work, in this work, the biomedical ontology problem is regarded as a discrete MOP, and a multiobjective SIA is presented to address it, which takes into consideration the DM's preference.

3. Preliminaries

3.1. Biomedical Ontology Matching Problem. Ontology matching aims at determining the identical entity mappings, which is the so-called ontology alignment. In the past, the ontology alignment's quality was often measured by f-measure [21], which is defined as follows:

$$\begin{aligned} recall &= \frac{|R \cap A|}{|R|}, \\ precision &= \frac{|R \cap A|}{|A|}, \\ f-measure &= \frac{recall \times precision}{\alpha \times recall + (1-\alpha) \times precision}, \end{aligned} \tag{1}$$

where R and A are, respectively, the alignments determined by the domain expert and ontology matcher. Recall and precision, respectively, measure A's completeness and soundness, f-measure is their harmony mean, and $\alpha \in [0, 1]$ is the weight to trade-off recall and precision. However, f-measure requires domain experts to provide R, which limits its application in real practice. Supposing a golden biomedical ontology alignment's cardinality is 1 : 1, since the larger cardinality of the concept mapping set is and the higher the found correspondences' mean similarity value is, the better the alignment quality is, three approximate measures, i.e., recall', precision', and f-measure', are used to, respectively, approximate the original recall, precision, and f-measure [22]:

$$\begin{aligned} recall' &= \frac{|A|}{\max\{|C_1|, |C_2|\}}, \\ precision' &= \frac{\sum_{i=1}^{|A|} simValue_i}{|A|}, \\ f-measure' &= \frac{recall' \times precision'}{\alpha \times recall' + (1-\alpha) \times precision'}, \end{aligned} \tag{2}$$

where $|C_1|$and $|C_1|$are two ontologies' concept scale, $|A|$is A's concept mapping number, and $simValue_i$is the i-th concept mapping's similarity value.

Finally, biomedical ontology matching problem's discrete multiobjective optimal model is defined as follows:

$$\begin{cases} \max f(x) = (r(x), p(x)), \\ s.t.\ X = \left(x_1, x_2, \ldots, x_{|C_1|}\right)^T, \\ x_i \in \{1, 2, \ldots, |C_2|\}, i = 1, 2, \cdots, |C_1|, \end{cases} \tag{3}$$

where $|C_1|$ and $|C_2|$ are, respectively, the concept number of two biomedical ontologies, $x_i$means the i-th pair of concept correspondence, i.e., i-th source concept is mapped to target x_i-th concept, and two objective functions r and p are to, respectively, maximize recall' and precision' of X's corresponding alignment.

3.2. Similarity Measure on Biomedical Concept. In this work, a hybrid similarity measure is used to enhance the confidence of the calculated similarity value, which takes into consideration the concept's syntax and linguistic and context information. First, for each biomedical concept, the information (the label, comment, and property name) from itself and its context concept are put into its separated profile. Then, two biomedical concepts' similarity value is calculated according to the following equation:

$$sim(p_1, p_2) = \frac{\sum_{i=1}^{f} \max_{j=1\cdots g}\left(sim\left(p_{1i}, p_{2j}\right)\right) + \sum_{j=1}^{g} \max_{i=1\cdots f}\left(sim\left(p_{1i}, p_{2j}\right)\right)}{f+g}, \tag{4}$$

where $p_1$and $p_2$are, respectively, two biomedical concept's corresponding profiles and $sim(p_{1i}, p_{2j})$ is calculated by integrating the SMOA [23] based on syntax measure and Unified Medical Language System (UMLS) [24] based on linguistic measure.

$$sim\left(p_{1i}, p_{2j}\right) = \begin{cases} 1, & \text{if two elements are synonymous in UMLS} \\ SMOA\left(p_{1i}, p_{2j}\right), & \text{otherwise} \end{cases}. \tag{5}$$

4. Compact Multiobjective Swarm Optimization Algorithm Driven by Knee Solution

This work proposes a CMPSO-K for solving the biomedical ontology matching problem, which approximates the population-based PSO's evolving process through a probability vector (PV) [25]. In the next, the objective decomposition approach, the encoding mechanism, and knee solution determination are presented, respectively, and finally, the details of CMPSO-K are shown through the pseudo-code.

4.1. Objective Decomposition. In this work, the weighted sum approach is used to transform a MOP into a set of subproblems and solves them simultaneously.

$$\begin{cases} \max g^{ws}(xw) = \sum_{i=1}^{N} w_i \cdot f_i(x) \\ s.t.\ x \in \Omega \subset R^n \end{cases}, \tag{6}$$

where N is the number of decomposed problems and $w_i \in [0, 1]\, i = 1, \ldots, N, \sum_{i=1}^{N} w_i = 1$. Two objectives of our problem are to maximize recall' and precision', and the ith subproblem's objective is defined as follows:

$$\max \frac{recall' \times precision'}{\alpha_i \times recall' + (1-\alpha_i) \times precision'}, \tag{7}$$

where $\alpha_1 = 0$ and $\alpha_i = i - 1/N - 1, \ \ldots \infty, \alpha_N = 1$.

4.2. Encoding Mechanism. In this work, the Gray coding, a binary encoding mechanism, is used to encode an alignment. As can be seen from Figure 1, an example of the encoding mechanism is shown, in which the source concept "uterine gland" with index 8 is mapped to target concept "Uterine Gland" with index 6 whose Gray code is 110. In particular, Gray code 000 means a source concept is not mapped to any target concept.

Moreover, a PV is utilized to characterize a population for solving one decomposed subproblem, whose element number is equal to that of an individual, and each element in PV represents the probability of being 0. Therefore, various individuals can be generated with the binary code through a PV. Figure 2 shows an example of generating an individual through PV. Through comparing each of PV's element with the random number, the new individual's element value can be determined. It is obvious that when all of PV's elements are close to 1 or 0, the algorithm tends to converge. At the end of each generation, PV is updated according to the best individual by increasing (or decreasing) its element value if the corresponding element of that individual is 1 (or 0).

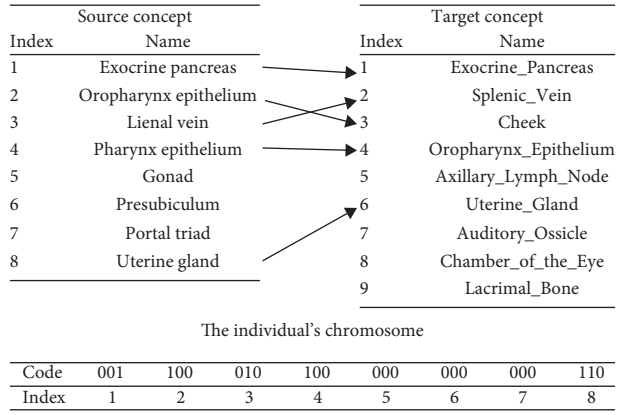


Figure 1: An example of the encoding mechanism.

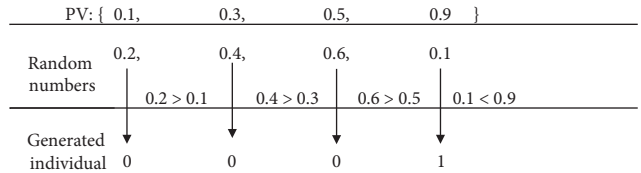


Figure 2: An example of generating an individual through PV.

Input: two individuals ind_1 and ind_2, crossover probability p_{cr}
Output: a new individual ind_{new}
(1) $ind_{\text{new}} = ind_1$;
(2) **for** $ind\ i = 1$; $i \le ind_1\ .length$, i++ **do**
(3) **if** $rand(0, 1) < p_{cr}$ **then**
(4) $ind_{\text{new},i} = ind_{2,i}$;
(5) **end if**
(6) **end for**
(7) return ind_{new};

Algorithm 1: Uniform crossover.

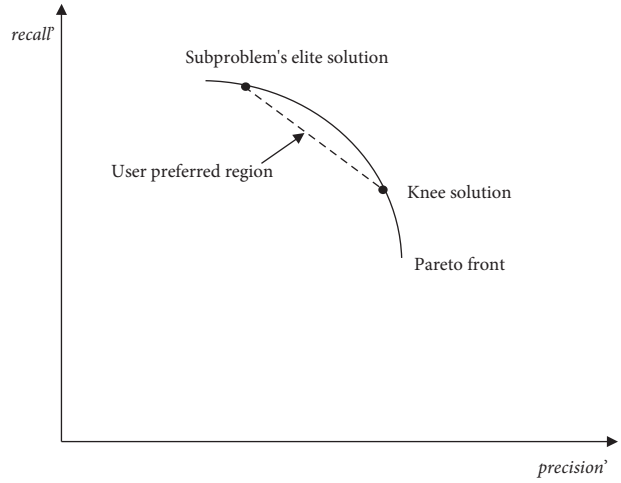


Figure 3: An example of knee solution and user preferred region.

Input: PF solution set $IND_{pf} = \{ind_1, ind_2, \cdots\}$
Output: a knee solution ind_{elite}
(1) max = 0;
(2) **for** ind $i = 1$; $i \leq IND_{pf}$.*length*, $i++$ do
(3) $\min_i = \min\{ind_i.recall', ind_i.precision'\}$
(4) **if** $\min_i >$ max **then**
(5) $\max = \min_i$;
(6) $ind_{elite} = ind_i$;
(7) **end if**
(8) **end for**
(9) return ind_{elite}

Algorithm 2: Determination of knee solution.

4.3. Crossover Operator. The crossover operator generates one child individual by mixing the information of two parent individuals, which is carried out according to the crossover probability. In this work, the uniform crossover operator is used, whose pseudo-code is shown in Algorithm 1. Given two parent solutions ind_1 and ind_2, each gene bit value of their offspring ind new is the same with the corresponding gene bit value of ind_1 (or ind_2) when a random number $rand(0, 1)$ is larger than (or smaller than) the crossover rate p_{cr}.

4.4. Determination of Knee Solution. Figure 3 shows an example of knee solution and user preferred region. As can be seen from the figure, the knee solution, a subproblem's elite solution, and the PF form a user preferred region for that subproblem. In each generation, the newly generated solution tends to move toward that region to meet a DM's requirement. Thus, the knee solution's determination is of utmost importance for the algorithm's search performance. In this work, for those solutions in the PF, a max-min approach is used to determine the knee solution. In particular, suppose ind_1 and ind_2 are two solutions in the PF, and $recall_i'$ and $precision_i'$, $i = 1, 2$, are, respectively, their recall' and precision', and they can be compared according to the following formula:

$$x_i = \arg_i \max\{\min\{recall_i', precision_i'\}\}. \quad (8)$$

On this basis, a best solution can be selected from PF solutions, and this procedure is shown in Algorithm 2.

4.5. Pseudo-Code of Compact Multiobjective Particle Swarm Optimization Algorithm Driven by Knee Solution. The pseudo-code of CMPSO-K is presented in Algorithm 3. CMPSO-K first divides the problem into three subproblems that, respectively, maximize $f_{\alpha=0}$, $f_{\alpha=0.5}$, and $f_{\alpha=1}$. First, three PVs and local best individuals are, respectively, initialized for three subproblems, and then the knee solution (or global best individual) is initialized by first using the nondominated sorting algorithm [26] on the population and then the knee solution in the PF is determined. In each generation, CMPSO-K tries to solve each subproblem by approximating PSO's position updating strategy, i.e., crossover an individual with the local best individual and global best individual to obtain a new individual, and then use the new one to update the local best individual and PV. After solving each subproblem, a population in the current generation is obtained by 3 local best individuals and 6 individuals in total generated through $PV_{\alpha=0}$, $PV_{\alpha=0.5}$, and $PV_{\alpha=1}$, respectively, and then the nondominated sorting algorithm is used to determine its PF and the current generation's knee solution ind'_{elite}. Through comparing ind'_{elite} with historical knee solution ind_{elite} through the max-min approach, the latter can be updated. Finally, when the generation approaches the maximum generation *maxGen*, the algorithm terminates and returns ind_{elite}.

5. Experiment

To test CMPSO-K's performance, the experiment exploits three biomedical tracks in Ontology Alignment Evaluation Initiative (OAEI), i.e., anatomy track, large biomed track, and disease and phenotype track. The testing cases in these tracks are all practical ontologies that are widely used in the biomedical domain and open to achieve, which have lots of overlapping information with different representations. OAEI provides the reference alignments for each track to test a matcher's performance, and Table 1 briefly describes the ontologies in these tracks.

In terms of the alignment's quality, the EA-based matcher [8], ABC-based matcher [6], PSO-based matcher [10], and OAEI's participants are compared in Tables 2 and 3, and also, in terms of the memory consumption and converging speed, CMPSO-K-based matcher and other SIA-based matchers are compared in Figures 4 and 5. All the SIAs' results are the mean values of 30 independent executions. EA, ABC, and PSO's configurations are referred to their literatures, and CMPSO-K uses the following configuration: number of decomposed problems: $N = 3$; maximum generation: $maxGen = 3000$; crossover probability: $p_{cr} = 0.6$; and step length for updating PV:$step = 0.1$. This configuration is determined in an empirical way, which represents a trade-off setting in the experiment to obtain the highest average results on all testing cases.

First, Friedman's test [32] is used to figure out whether all the matchers present any difference and then determines whether one matcher statistically outperforms others through Holm's test [33]. In Friedman's test, the null

Input: maximum generation $maxGen$, crossover rate p_{cr}, step length for updating PV $step$
Output: knee solution ind_{elite}

(1) **Initialization**
(2) initialize generation $t = 0$;
(3) initialize $PV_{\alpha=0}$, $PV_{\alpha=0.5}$ and $PV_{\alpha=1}$ by setting all the elements inside as 0.5;
(4) initialize three local best individuals $ind_{\alpha=0,elite}$, $ind_{\alpha=0.5,elite}$ and $ind_{\alpha=1,elite}$;
(5) generate 5 individuals through $PV_{\alpha=0}$, $PV_{\alpha=0.5}$ and $PV_{\alpha=1}$;
(6) $NonDominatedSort()$;
(7) initialize the knee solution (or global best individual) ind_{elite};
(8) **Evolving Process**
(9) **while** $t < maxGen$ **do**
(10) **Updating $ind_{\alpha=0,elite}$**
(11) generate an individual $ind_{\alpha=0,new}$ through $PV_{\alpha=0}$;
(12) $ind_{\alpha=0,new} = crossover(ind_{\alpha=0,new}, ind_{\alpha=0,elite})$
(13) $ind_{\alpha=0,new} = crossover(ind_{\alpha=0,new}, ind_{elite})$;
(14) $[winner, loser] = compete(ind_{\alpha=0,new}, ind_{\alpha=0,elite})$;
(15) $updatePV(winner, loser)$;
(16) **if** $winner == ind_{\alpha=0,new}$ **then**
(17) $ind_{\alpha=0,elite} = ind_{\alpha=0,new}$;
(18) **end if**
(19) **Updating $ind_{\alpha=0.5,elite}$**
(20) generate an individual $ind_{\alpha=0,new}$ through $PV_{\alpha=0.5}$;
(21) $ind_{\alpha=0.5,new} = crossover(ind_{\alpha=0,new}, ind_{\alpha=0.5,elite})$;
(22) $ind_{\alpha=0.5,new} = crossover(ind_{\alpha=0,new}, ind_{elite})$;
(23) $[winner, loser] = compete(ind_{\alpha=0,new}, ind_{\alpha=0.5,elite})$
(24) $updatePV(winner, loser)$;
(25) **if** winner $== ind_{\alpha=0,new}$ **then**
(26) $ind_{\alpha=0.5,elite} = ind_{\alpha=0,new}$;
(27) **end if**
(28) **Updating $ind_{\alpha=1,elite}$**
(29) generate an individual $ind_{\alpha=1,new}$ through $PV_{\alpha=1}$;
(30) $ind_{\alpha=1,new} = crossover(ind_{\alpha=1,new}, ind_{\alpha=1,elite})$;
(31) $ind_{\alpha=1,new} = crossover(ind_{\alpha=1,new}, ind_{elite})$;
(32) $[winner, loser] = compete(ind_{\alpha=1,new}, ind_{\alpha=1,elite})$
(33) $updatePV(winner, loser)$;
(34) **if** $winner == ind_{\alpha=1,new}$ **then**
(35) $ind_{\alpha=1,elite} = ind_{\alpha=1,new}$;
(36) **end if**
(37) **Updating ind_{elite}**
(38) generate 2 individuals through $PV_{\alpha=0}$, $PV_{\alpha=0.5}$ and $PV_{\alpha=1}$, respectively,
(39) $NonDominatedSort$;
(40) determine current generation's knee solution ind'_{elite};
(41) $[winner, loser] = compete(ind_{elite}, ind'_{elite})$;
(42) **if** $winner == ind'_{elite}$ **then**
(43) $ind_{elite} = ind'_{elite}$;
(44) **end if**
(45) $t = t + 1$;
(46) **end while**
(47) **return** ind_{elite};

Algorithm 3: Compact multiobjective particle swarm optimization algorithm driven by knee solution.

Table 1: Description on the OAEI's tracks.

Track	Subtask	Ontology	Scale
Anatomy	MA-HA	Adult Mouse Anatomy (MA)	2,744 classes
		Human Anatomy (HA)	3,304 classes
Large biomed	FMA-NCI	Foundation Model of Anatomy (FMA)	78,989 classes
	FMA-SNOMED	Systemized Nomenclature of Medicine (SNOMED)	122,464 classes
	SNOMED-NCI	National Cancer Institute Thesaurus (NCI)	66,724 classes
Disease and phenotype	HP-MP	Human Phenotype Ontology (HP)	33,205 classes
		Mammalian Phenotype Ontology (MP)	32,298 classes
	DOID-ORDO	Human Disease Ontology (DOID)	24,034 classes
		Orphanet Rare Disease Ontology (ORDO)	68,009 classes

Table 2: Friedman's test on the alignment's quality. Each value represents the f-measure, and the number in round parentheses is the corresponding computed rank.

	Anatomy	FMA-NCI	FMA-SNOMED	NCI-SNOMED	HP-MP	DOID-ORDO	Average
AML [27]	0.94 (1.5)	0.93 (1.5)	0.83 (2)	0.80 (2)	0.84 (3)	0.64 (7)	0.83 (2.83)
LogMap [28]	0.89 (4)	0.92 (3)	0.79 (3)	0.77 (3)	0.85 (2)	0.84 (2)	0.84 (2.83)
XMap [29]	0.89 (4)	0.86 (5.5)	0.77 (4)	0.69 (4)	0.47 (8.5)	0.70 (6)	0.73 (2.33)
DOME [30]	0.76 (6)	0.86 (5.5)	0.33 (9)	0.64 (6)	0.47 (8.5)	0.60 (9)	0.61 (7.33)
POMAP++ [31]	0.89 (4)	0.88 (4)	0.40 (8)	0.68 (5)	0.68 (6)	0.83 (3)	0.73 (5.00)
EA [8]	0.68 (9)	0.73 (8)	0.64 (6)	0.53 (8)	0.70 (5)	0.61 (8)	0.65 (7.33)
ABC [6]	0.72 (7)	0.75 (7)	0.64 (6)	0.62 (7)	0.62 (7)	0.72 (5)	0.68 (6.50)
PSO [10]	0.70 (8)	0.72 (9)	0.64 (6)	0.51 (9)	0.72 (4)	0.75 (4)	0.67 (6.67)
CMPSO-K	0.94 (1.5)	0.93 (1.5)	0.85 (1)	0.82 (1)	0.86 (1)	0.86 (1)	0.88 (1.17)

Table 3: Holm's test on the alignment's quality.

i	Approach	z value	Unadjusted p value	$\alpha/k-i, \alpha = 0.05$
8	XMap	2.33	0.045	0.050
7	AML	2.83	0.004	0.025
6	LogMap	2.83	0.004	0.016
5	POMAP++	5.00	1.28×10^{-7}	0.012
4	ABC	6.50	9.82×10^{-8}	0.010
3	PSO	6.67	3.79×10^{-8}	0.008
2	EA	7.33	5.38×10^{-11}	0.007
1	DOME	7.33	5.38×10^{-11}	0.006

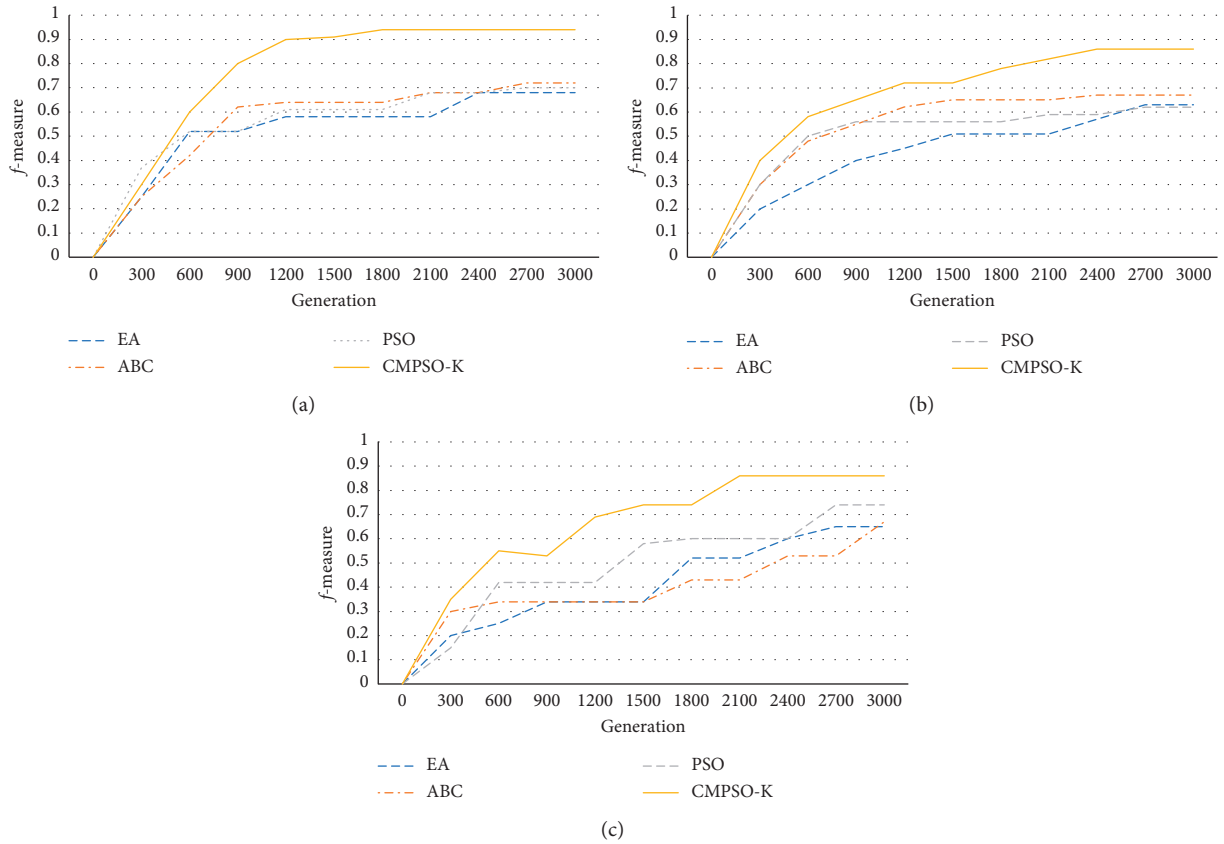


Figure 4: Comparison on the converging speed, (a) comparison on the anatomy track, (b) comparison on the large biomed track, and (c) comparison on the disease and phenotype track.

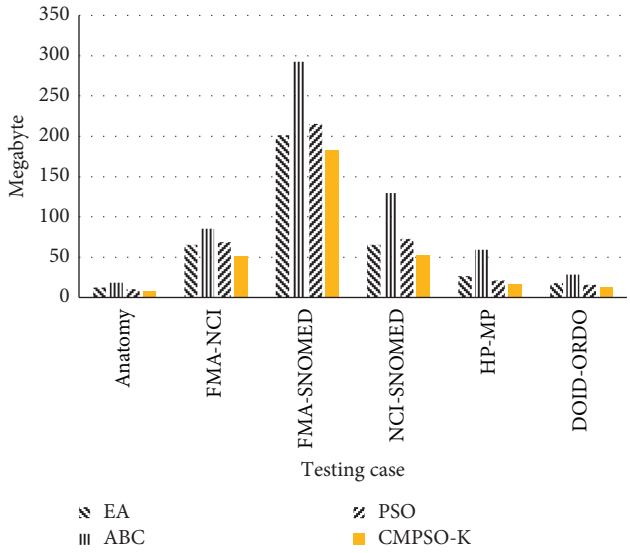


Figure 5: Comparison on the memory consumption.

hypothesis is that all the matchers are equivalent, and if the computed value χ_r^2 must be equal to or greater than the tabled critical chi-square value at the specified level of significance [34], this hypothesis can be rejected. In this work, a level of significance = 0.05 is chosen, and the critical value for 8 degrees of freedom (since there are 9 matchers), i.e., $\chi_{0.05}^2 = 15.507$, needs to be considered.

In Table 2, computed $\chi_r^2 = 54.43$, which is greater than 15.507, and therefore, the null hypothesis can be rejected, and Holm's test can be further carried out. As shown in Table 2, since the CMPSO-K-based matcher ranks with the lowest value, it is set as a control matcher that will be compared with others.

In Holm's test, z value is the testing statistic for comparing the i-th and j-th matchers, which is used for finding the p value that is the corresponding probability from the table of the normal distribution. p value is then compared with $\alpha = 0.05$, which is an appropriate level of significance. According to Table 3, it is possible to state that our approach statistically outperforms other biomedical ontology matchers on f-measure at 5% significance level. Since the multiobjective evolving mechanism can better trade-off two objectives and the knee solution can effectively guide the algorithm's search direction, CMPSO-K's solutions are much better than other SIAs.

Figures 4 and 5, respectively, compare CMPSO-K with other SIAs on the memory consumption and converging speed. As can be seen from the figures, CMPSO-K can significantly improve the converging speed and reduce the memory consumption, which shows the effectiveness of the compact encoding mechanism and the compact evolutionary operators. To sum up, CMPSO-K-based ontology matching technique can efficiently optimize the biomedical ontology alignments.

6. Conclusion and Future Work

To optimize biomedical ontology alignment's quality, in this paper, a discrete multiobjective optimal model is constructed, and a hybrid similarity metric to distinguish the heterogeneous biomedical concepts, a CMPSO-K-based ontology matching technique is then proposed for addressing it. Compared with the most existing SIA-based ontology matching techniques, CMPSO-K takes into consideration both the algorithm's performance and the DM's preference. Accordingly, three methods are proposed to achieve this goal: (1) the compact encoding mechanism for saving the memory consumption and runtime; (2) multiobjective decomposition and evolutionary mechanism for trading-off different objectives; and (3) the max-min strategy for determining the knee solution and guiding the algorithm's searching direction. Our work presents a novel compact multiobjective evolutionary framework that can improve the efficiency of the current SIA-based ontology matching technique.

In the experiment, the f-measure values obtained by CMPSO-K outperform all the other competitors, which shows that CMPSO-K can effectively optimize the ontology alignments. In particular, the quality of alignment of CMPSO-K is better than EA ABC and PSO, which shows that CMPSO-K's multiobjective evolutionary mechanism driven by the knee solution can effectively trade-off two optimal objectives and find the better solution. Moreover, since the compact encoding mechanism uses PVs to represent the swarms and the compact evolutionary operators can simplify the population-based PSO's evolving process, CMPSO-K can significantly reduce the memory consumption and runtime. Since none of the similarity measures can effectively distinguish all the heterogeneous concepts in any situations, it is necessary to aggregate several similarity measures to improve the result's precision. We utilize a hybrid similarity measure which combines three kinds of similarity measures to calculate the entity similarity value, and therefore, CMPSO-K's results are significantly higher than other systems that only take into consideration one or two categories of similarity measure, such as DOME, POMAP++, and LogMap. However, AML applies too many similarity measures that lead to the conflicting results, which decrease its recall value. Thus, how many similarity measures should be selected and combined to ensure the quality of the alignment will be one of our future work. Last but not the least, CMPSO-K takes into consideration of DM's preference and utilizes the knee solution to guide the algorithm's search direction, which can effectively trade-off two optimal objectives and find better solutions.

To further improve the matching process's efficiency, in the future, we will be interested in developing a biomedical ontology partitioning technique to split two ontologies into disjoint ontology segments so that the large-scale problem can be converted into several small-scale segment matching problems. After that, the parallel computation can also be further utilized to match the similar segments and improve the matching process's performance. With respect to the similarity measure, we would like to develop an adaptive biomedical concept similarity measure framework, which can instantiate an effective similarity metric according to two biomedical ontologies' heterogeneous characteristics.

Data Availability

The data used to support this study can be found in http://oaei.ontologymatching.org.

Conflicts of Interest

The authors declare that they have no conflicts of interest in the work.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (no. 61503082), the Natural Science Foundation of Fujian Province (no. 2016J05145), the Program for New Century Excellent Talents in Fujian Province University (no. GY-Z18155), the Program for Outstanding Young Scientific Researcher in Fujian Province University (no. GY-Z160149), and the Scientific Research Foundation of Fujian University of Technology (nos. GY-Z17162 and GY-Z15007).

References

[1] E. Zitzler and L. Thiele, "Multiobjective evolutionary algorithms: a comparative case study and the strength pareto approach," *IEEE Transactions on Evolutionary Computation*, vol. 3, no. 4, pp. 257–271, 1999.

[2] S. Sudeng and N. Wattanapongsakorn, "A knee-based multiobjective evolutionary algorithm: an extension to network system optimization design problem," *Cluster Computing*, vol. 19, no. 1, pp. 411–425, 2016.

[3] S. Sudeng and N. Wattanapongsakorn, "A decomposition-based approach for knee solution approximation in multiobjective optimization," in *Proceedings of the 2016 IEEE Congress on Evolutionary Computation (CEC)*, pp. 3710–3717, Vancouver, Canada, July 2016.

[4] G. Acampora, V. Loia, and A. Vitiello, "Enhancing ontology alignment through a memetic aggregation of similarity measures," *Information Sciences*, vol. 250, pp. 1–20, 2013.

[5] A. Alves, K. Revoredo, and F. Baião, "Ontology alignment based on instances using hybrid genetic algorithm," in *Proceedings of the 7th International Conference on Ontology Matching*, pp. 242-243, Rio de Janeiro, Brazil, November 2012.

[6] Y. He, X. Xue, and S. Zhang, "Using artificial bee colony algorithm for optimizing ontology alignment," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 8, no. 4, pp. 766–773, 2017.

[7] J. Martinez-Gil and J. F. Aldana-Montes, "Evaluation of two heuristic approaches to solve the ontology meta-matching problem," *Knowledge and Information Systems*, vol. 26, no. 2, pp. 225–247, 2011.

[8] J. Wang, Z. Ding, and C. Jiang, "Gaom: Genetic algorithm based ontology matching," in *Proceedings of the IEEE Asia-Pacific Conference on Services Computing (APSCC'06)*, pp. 617–620, Guangzhou, China, December 2006.

[9] X. Xue and Y. Wang, "Improving the efficiency of NSGA-II based ontology aligning technology," *Data & Knowledge Engineering*, vol. 108, pp. 1–14, 2017.

[10] J. Bock and J. Hettenhausen, "Discrete particle swarm optimisation for ontology alignment," *Information Sciences*, vol. 192, pp. 152–173, 2012.

[11] W. Z. Guo, J. Y. Chen, G. L. Chen, and H. F. Zheng, "Trust dynamic task allocation algorithm with Nash equilibrium for heterogeneous wireless sensor network," *Security and Communication Networks*, vol. 8, no. 10, pp. 1865–1877, 2015.

[12] G. Liu, W. Guo, R. Li, Y. Niu, and G. Chen, "XGRouter: high-quality global router in X-architecture with particle swarm optimization," *Frontiers of Computer Science*, vol. 9, no. 4, pp. 576–594, 2015.

[13] W. Guo, J. Li, and G. Chen, "A PSO-optimized real-time fault-tolerant task allocation algorithm in wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 12, pp. 3236–3249, 2014.

[14] H. Guo, B. Liu, D. Cai, and T. Lu, "Predicting protein-protein interaction sites using modified support vector machine," *International Journal of Machine Learning and Cybernetics*, vol. 9, no. 3, pp. 393–398, 2018.

[15] G. Alexandru-Lucian and A. Iftene, "Using a genetic algorithm for optimizing the similarity aggregation step in the process of ontology alignment," in *Proceedings of the 9th Roedunet International Conference*, pp. 118–122, Sibiu, Romania, June 2010.

[16] X. Xue and Y. Wang, "Optimizing ontology alignments through a memetic algorithm using both matchfmeasure and unanimous improvement ratio," *Artificial Intelligence*, vol. 223, pp. 65–81, 2015.

[17] X. Xue, J. Lu, and J. Chen, "Using NSGA-III for optimising biomedical ontology alignment," *CAAI Transactions on Intelligence Technology*, vol. 4, no. 3, pp. 135–141, 2019.

[18] H. Seada and K. Deb, "U-NSGA-III: a unified evolutionary optimization procedure for single, multiple, and many objectives: proof-of-principle results," in *Proceedings of the International conference on evolutionary multi-criterion optimization*, pp. 34–49, Guimarães, Portugal, March 2015.

[19] S. C. Chu, X. Xue, J. S. Pan, and X. Wu, "Optimizing ontology alignment in vector space," *Journal of Internet Technology*, vol. 21, no. 1, pp. 15–23, 2020.

[20] X. Xue, "Compact firefly algorithm for matching biomedical ontologies," *Knowledge and Information Systems*, 2020.

[21] C. J. Van Rijsbergen, "Foundation of evaluation," *Journal of Documentation*, vol. 30, no. 4, pp. 365–373, 1974.

[22] X. Xue, J. Chen, and X. Yao, "Efficient user involvement in semi-automatic ontology matching," *IEEE Transactions on Emerging Topics in Computational Intelligence*, pp. 1–11, 2019.

[23] G. Stoilos, G. Stamou, and S. Kollias, "A string metric for ontology alignment," in *4th International Semantic Web Conference (ISWC 2005)*, pp. 623–637, Springer, Berlin, Germany, 2005.

[24] O. Bodenreider, "The unified medical language system (umls): integrating biomedical terminology," *Nucleic Acids Research*, vol. 32pp. D267–D270, suppl_1, 2004.

[25] G. R. Harik, F. G. Lobo, and D. E. Goldberg, "The compact genetic algorithm," *IEEE Transactions on Evolutionary Computation*, vol. 3, no. 4, pp. 287–297, 1999.

[26] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.

[27] D. Faria, C. Pesquita, T. Tervo, F. M. Couto, and I. F. Cruz, "AML and AMLC results for OAEI 2019," in *Proceedings of the 18th International Semantic Web Conference (ISWC 2019)*, pp. 101–106, Auckland, New Zealand, October 2019.

[28] E. Jimenez-Ruiz, "LogMap family participation in the OAEI 2019," in *Proceedings of the 18th International Semantic Web*

Conference (ISWC 2019), pp. 160–163, London, UK, January 2019.

[29] W. E. Djeddi and M. T. Khadir, "XMAP: a novel structural approach for alignment of OWL-full ontologies," in *Proceedings of the 2010 International Conference on Machine and Web Intelligence*, pp. 368–373, Algiers, Algeria, October 2010.

[30] S. Hertling and H. Paulheim, "DOME results for OAEI 2019," in *Proceedings of the 18th International Semantic Web Conference (ISWC 2019)*, pp. 123–130, Auckland, New Zealand, October 2019.

[31] A. Laadhar, F. Ghozzi, I. Megdiche, F. Ravat, O. Teste, and F. Gargouri, "Pomap++ results for OAEI 2019: fully automated machine learning approach for ontology matching," in *Proceedings of the 18th International Semantic Web Conference (ISWC 2019)*, pp. 169–174, Auckland, New Zealand, October 2019.

[32] M. Friedman, "The use of ranks to avoid the assumption of normality implicit in the analysis of variance," *Journal of the American Statistical Association*, vol. 32, no. 200, pp. 675–701, 1937.

[33] S. Holm, "A simple sequentially rejective multiple test procedure," *Scandinavian Journal of Statistics*, vol. 6, no. 2, pp. 65–70, 1979.

[34] M. S. Nash, "Handbook of parametric and nonparametric statistical procedures," *Technometrics*, vol. 43, no. 3, p. 374, 2001.