An effectiveness model of vocational education mode reform based on data mining

Jiaen Gu

Zibo Vocational Institute,
Zibo 255314, China
Email: jiaen@36haojie.com

Abstract: The existing education mode reform model is not deep enough in data mining and there are problems of poor accuracy and low credibility. This paper proposes the construction of the effectiveness model of vocational education mode reform based on data mining. This paper also analyses the effective indicators of vocational education mode reform, obtains the indicators of vocational education mode reform; uses the Simhash algorithm to remove and cleans the indicators of vocational education mode reform, and uses the data information gain algorithm to select the characteristics of indicators. It also constructs the topological structure of the effectiveness model, determines the input and output parameters of the model, introduces the data mining method to mine the effectiveness index data, and completes the construction of the effectiveness model of vocational education mode reform. The experimental results show that the validity index of the model is 26.79, and the analysis accuracy of the model can reach 97.2%.

Keywords: data mining; vocational education; mode reform; effectiveness model.

Reference to this paper should be made as follows: Gu, J. (2022) 'An effectiveness model of vocational education mode reform based on data mining', *Int. J. Continuing Engineering Education and Life-Long Learning*, Vol. 32, No. 1, pp.111–127

Biographical notes: Jiaen Gu received his Master's in Modern Chinese History from Guangxi Normal University in 2006. He is currently a Lecturer of Marxism College of Zibo Vocational Institute. His research interests include history, ideological and political education, and vocational education theory.

1 Introduction

In recent years, with the deepening of structural reform in China, the demand for professional talents is also increasing, which leads to the fact that the existing education mode has been unable to adapt to the current social development. For the field of education, vocational education is mainly to cultivate professional talents with professional knowledge, skills and ethics required by a certain profession or production labour, which meets the current talent demand of China. Vocational education mainly includes vocational school education and vocational training. Vocational school education refers to vocational and technical schools, technical schools, etc., belonging to education for academic qualifications, which are divided into primary, secondary and higher vocational school education; vocational training refers to pre-employment training and retraining for employees. Compared with general education, vocational education aims to cultivate application-oriented talents with certain cultural level and professional skills, with more emphasis on the cultivation of practical work ability and practical skills. Vocational education is the inevitable outcome of social development, which benefits each other and promotes the rapid development of society. Vocational education mode refers to the specific pattern formed under certain social conditions (Kim, 2018; Ashby and Maher, 2018; Adkins et al., 2017). The current vocational education mode in China has the characteristics of unification, centralisation, closure and mechanisation, which leads to the rigid content and methods of vocational education and the inability to cultivate professional talents for social needs. Therefore, it is of practical significance to reform the vocational education mode and adapt to social needs. In order to analyse the role of the reform of vocational education mode, a model is built to verify the effectiveness of the reform of vocational education mode (Finge, 2017; Helma et al., 2018; Popovic, 2017).

Many scholars have studied the effectiveness model of education mode reform. Chen and Wang (2018) propose the effectiveness model of vocational education mode reform based on ant colony algorithm, which mainly estimates the effective probability of vocational education mode reform according to ant colony algorithm. The algorithm has the characteristics of positive information feedback, and can accurately estimate the effectiveness of vocational education mode reform, so as to realise the effectiveness analysis of the vocational education mode reform. This method can effectively evaluate the effectiveness of education model reform and has certain credibility. However, the analysis of this method takes a long time, which is not conducive to wide application. Xue and Sha (2017) put forward an effective model of vocational education mode reform based on clustering algorithm, which makes statistics and analysis on the data of vocational education mode reform according to clustering algorithm. On this basis, an effective analysis model is constructed to calculate the effectiveness of vocational education mode reform. This method effectively clusters the reform data of vocational education mode, and realises the construction of the reform effectiveness model of vocational education mode. However, the validity index analysis of this method is not accurate. Hu et al. (2018) put forward the effectiveness model of vocational education mode reform based on the random forest algorithm, which mainly uses the random forest algorithm to classify the reform data of vocational education mode, and constructs the corresponding effectiveness model of vocational education mode reform based on the classification results. This method uses the random forest algorithm to classify the vocational education model reform data, which can be evaluated for a certain category. However, the anti-interference ability of the model is poor in the overall application process, and the analysis accuracy is poor. The model can not accurately analyse the effectiveness of vocational education model reform.

Each of the above three models has its own advantages, but there are problems of low model accuracy and effectiveness indicator, which cannot meet the analysis needs of the effectiveness of the current vocational education mode reform. Therefore, the effectiveness model of vocational education mode reform based on data mining is proposed to achieve the accurate research of the effectiveness of vocational education mode reform. The technical route of this paper is as follows:

1 This paper analyses the effectiveness of vocational education model reform, as well as the effective indicators of vocational education model reform. To prepare for the construction of the effectiveness analysis model of vocational education model reform.

2 Using the Simhash algorithm to de duplicate and clean the effective index data provides accurate data support for the selection of effective index features.

3 The data information gain algorithm is used to select the characteristics of effective indicators, which lays the foundation for the construction of the effectiveness analysis model of vocational education model reform.

4 Experiments are carried out to verify the validity of the model.

5 Conclusion and future development direction.

2 The effectiveness model of vocational education mode reform based on data mining

2.1 Analyses of effectiveness indicators of vocational education mode reform

In order to build an accurate analysis model of the effectiveness of vocational education mode reform, the first task is to analyse the effectiveness indicators of vocational education mode reform, and prepare for the construction of the effectiveness analysis model of vocational education mode reform. In order to analyse the effectiveness of vocational education mode reform and the effectiveness indicators of vocational education mode reform, the currently widely used effectiveness indicators of vocational education mode reform (Teixeira and Koryakina, 2016) are shown in Table 1.

As shown in Table 1, it is the effectiveness indicator of the current vocational education mode reform. In order to facilitate the construction of the effectiveness analysis model of the vocational education mode reform, the above effectiveness indicators are normalised and recorded as $R_i = \{r_1, r_2, \ldots, r_n\}$.

Through the above process, the analysis of the effectiveness indicators of vocational education mode reform is completed, which makes preparation for the construction of the effectiveness analysis model of vocational education mode reform.

2.2 De duplication and cleaning of effectiveness indicator data

Based on the above analysis of the effectiveness indicators of vocational education mode reform, the effectiveness indicator data of vocational education mode reform is obtained, and the Simhash algorithm is used to de duplicate and clean the effectiveness indicator data. The specific treatment process of effectiveness indicators is as follows.

High quality effectiveness indicator data is the foundation of the effectiveness analysis model of vocational education mode reform, so it is necessary to de duplicate and clean the obtained effectiveness indicator data.

If the effectiveness indicator data is repeated, it will not only increase the pressure of effectiveness analysis and calculation, but also lead to the deviation of effectiveness analysis results. In order to reduce the adverse effect of effectiveness indicator data on subsequent operations, the Simhash algorithm is used to de duplicate the effectiveness indicator data, and the implementation flow chart of effectiveness indicator data de duplication is shown in Figure 1.

Table 1 Effectiveness indicators of vocational education mode reform

First level indicator	*Two level indicator*	*Indicator significance*
Vocational education investment	Proportion of vocational education funds to government financial revenue	Development of vocational education
	The proportion of vocational education funds to GNP	
	Ratio of vocational education funds to average funds	
	Achievement rate of hardware facilities in vocational education	
The development speed of vocational education	Annual growth rate of vocational education funds	Development of vocational education
	Annual growth rate of school enrolment	
	Average growth rate of the number of students in school	
Education level for all	Average length of education	The achievement of the goals of vocational education
	Adult literacy rate	
Efficiency and benefit of vocational education	Proportion of teachers and students in vocational education schools	The achievement of the goals of vocational education
	Dropout rate in vocational education schools	
	The contribution rate of vocational education to regional economic development	
	The rate of students' quality reaching the standard	
The information level of vocational education	The popularisation rate of modern educational means	The development degree of vocational education modernisation
	Opening rate of computer course	

In the process of effectiveness indicator data de duplication, in order to ensure the best de duplication effect, considering the size of effectiveness indicator data set, the number of digits of the Simhash algorithm is set to 32, and the Simhash value of effectiveness indicator data is calculated. Based on this, the similarity of effectiveness indicator data is determined. Where, the calculation formula of the Simhash value is

$$\omega_{R_i} = \frac{\sum_{i=1}^{n} R_i}{\alpha^2 \Theta \gamma} \tag{1}$$

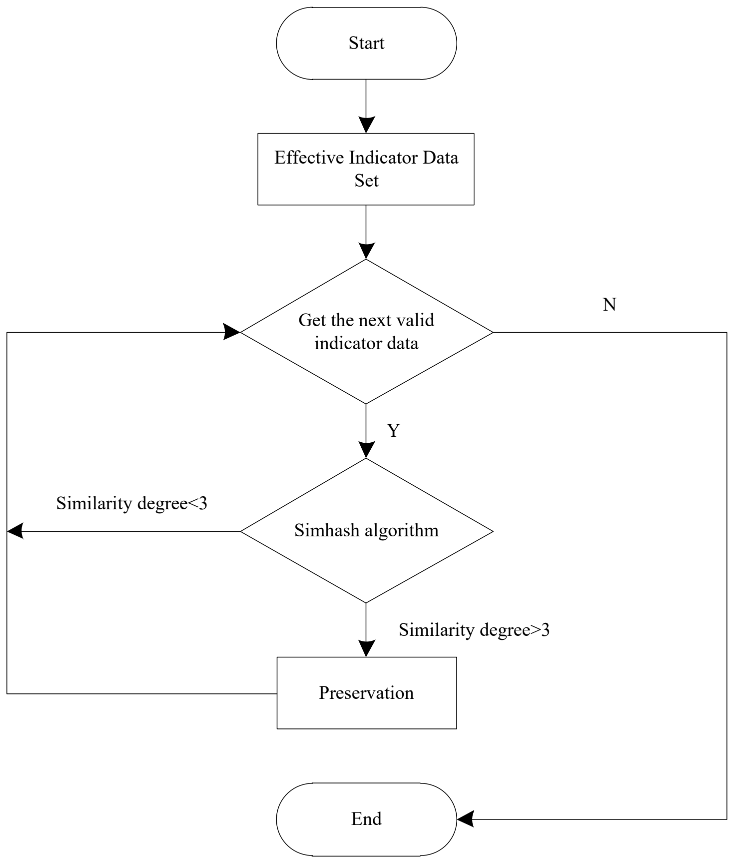
where ω_{R_i} represents the Simhash value of R_i; α represents the auxiliary parameter for calculating the Simhash value; γ represents the digits of the Simhash algorithm.

Then the de duplication rule of effectiveness indicator data is

$$\begin{cases} \omega_{R_i} \geq \omega^* & \text{Preserving valid indicator data} \\ \omega_{R_i} \leq \omega^* & \text{Delete valid indicator data} \end{cases} \tag{2}$$

where ω^* represents the critical value of determining similarity. According to the characteristics of the acquired effectiveness indicator data, the critical value of determining similarity is set as three.

Figure 1 Flow chart of effectiveness indicator data de duplication



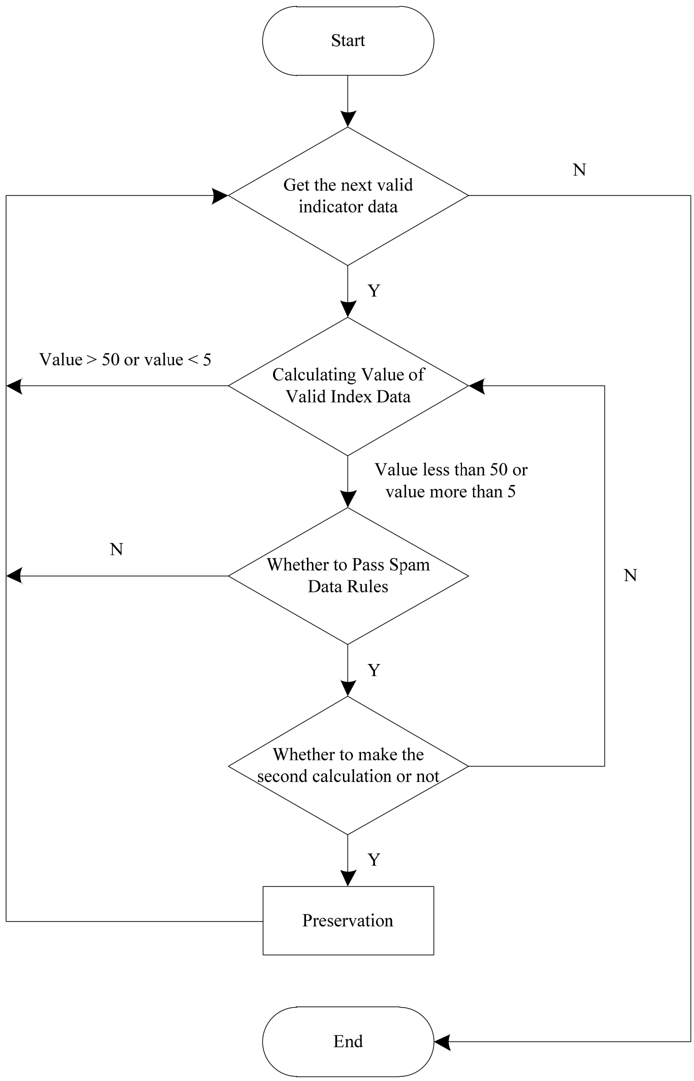
Through the above process, the de duplication of effectiveness indicator data is completed, but due to the different acquisition background, environment and other factors, the value of the effectiveness indicator data obtained is quite different. Therefore, there are a lot of redundant data in the acquired effectiveness indicator data, which will not only have a great adverse impact on the accuracy of the effectiveness indicator data, but also cause the deviation of effectiveness analysis, therefore, it needs to be cleaned . In order to ensure the cleaning effect of the effectiveness indicator data, the garbage data in the de duplicated effectiveness indicator data is processed firstly, and the garbage data identification formula is

$$r_i' = \sum_{i=1}^{n} \frac{R_i \beth}{\beta^3} \tag{3}$$

where r_i' refers to garbage data; $\beth$ refers to the parameter of garbage data identification; β refers to auxiliary factor.

The flow chart of effectiveness indicator data cleaning is shown in Figure 2.

Figure 2 Flow chart of effectiveness indicator data cleaning



As shown in Figure 2, the calculation formula of effectiveness indicator data value is

$$V_{R_i} = \sum_{i=1}^{n} R_i * \prod \Phi\Theta\Psi \tag{4}$$

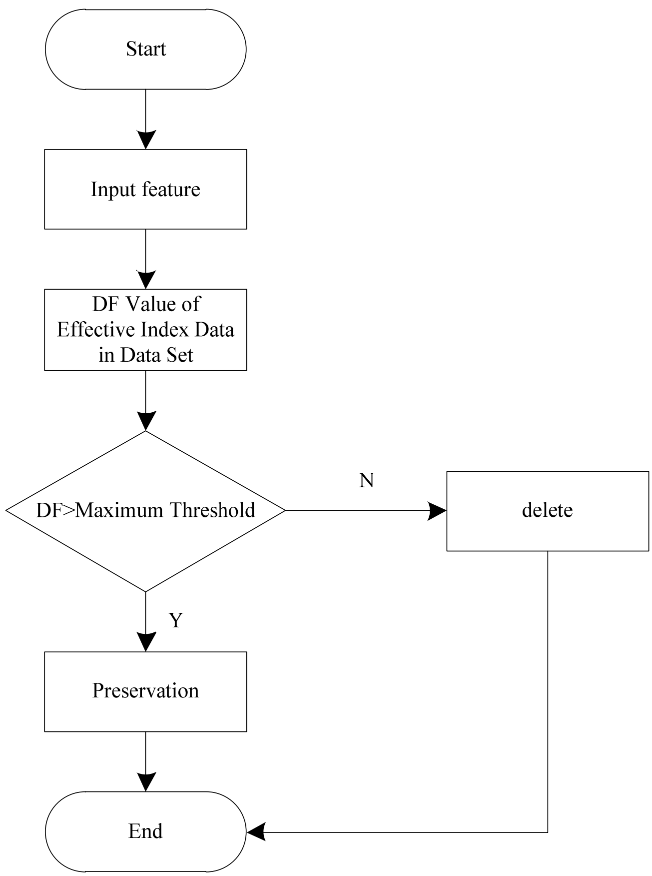
where V_{R_i} represents the value of effectiveness indicator data, $\prod \Phi\Theta\Psi$ is the auxiliary parameter of value calculation, and the value of Φ and Ψ is mainly determined by the size of the effectiveness indicator data set.

Through the above process, the de duplication and cleaning of effectiveness indicator data are completed, providing accurate data support for the selection of the following effectiveness indicator characteristics (Chong and Yong, 2016).

2.3 Selections of effectiveness indicator characteristics

Based on the effectiveness indicator data processed above, the data information gain algorithm is used to select the effectiveness indicator characteristics (Bini et al., 2017). The specific selection process is as follows.

Figure 3 Calculation flow chart of the frequency characteristics of effectiveness data



The frequency characteristics of effectiveness data refer to the frequency of effectiveness data acquisition; the weight of effectiveness data mainly represents the importance of effectiveness indicator data; data information gain refers to the amount of effectiveness indicator data information. In general, the larger the data information gain is, the more important the characteristics of effectiveness indicator data are (Gori et al., 2017; Thibault, 2015).

Where, the calculation flow of the frequency characteristics of effectiveness data is shown in Figure 3.

The calculation formula of effectiveness data weight is

$$Q_i = \frac{\omega_{R_i - r_i'}}{v_{R_i}} \cdot \int \vartheta^4 \tag{5}$$

where $\int \vartheta^4$ is the weight calculation factor.

The calculation process of data information gain is complex. Firstly, for the effectiveness indicator data, there are n possibilities of data information gain, which are recorded as $c_1, c_2, \ldots, c_n$. The corresponding probability is expressed as $p_1, p_2, \ldots, p_n$, then the definition formula of data information gain is

$$H(c) = -\sum_i^n p_i * \log p_i \tag{6}$$

Then the expression of information gain of effectiveness indicator data is

$$IG(R_i) = H(c) - H\left(\frac{c}{R_i}\right) \tag{7}$$

where $H\left(\frac{c}{R_i}\right)$ is the conditional entropy, and its expansion is expressed as

$$H\left(\frac{c}{R_i}\right) - \sum_i p\left(\frac{c}{R_i}\right) * \left(\frac{c}{R_i}\right) \tag{8}$$

According to the above formula, the final expression of the information gain of effectiveness indicator data is

$$IG(T) = Q_i - \sum_i^n p_i * \log p_i - \sum_i^n p\left(\frac{c}{R_i}\right) * \log\left(\frac{c}{R_i}\right) \tag{9}$$

Because the extraction of effective index features is related to the correlation coefficient. Therefore, the concept of association coefficient is introduced into data mining. The characteristics of association data are relatively scattered, which is not convenient for mining as a whole. At this time, it is necessary to centralise the association information, and centralise the processing through the method of average value. The calculation formula is:

$$\phi_0 = \frac{1}{r} \sum_{k=1}^r \xi_{ik} \tag{10}$$

If the index features are given different weights, the calculation formula is

$$\phi_0 = \frac{1}{r} \sum_{k=1}^r w \xi_{ik} \tag{11}$$

where w_k is the weight of the first indicator.

Through the processing of the above association information weight, the effective index feature mining is realised.

In the effective data frequency feature extraction algorithm, this paper proportions the effective data weight, analyses the multiple possibilities of data information gain, distinguishes the multiple possibilities, introduces the correlation coefficient, and obtains the calculation result of the effective index data information gain. This result can be used as the final target of the effective index feature extraction in this data mining.

According to the above process, the selection of effectiveness indicator characteristics is completed, which is ready for the construction of the effectiveness analysis model of the final vocational education mode reform (Gambarotto and Oken, 2017).

2.4 Construction of effectiveness model based on data mining

In order to improve the accuracy of the effectiveness model, we select the data of 30 vocational colleges' education mode reform since 2015, and build the effectiveness model based on the data.

This paper analyses the research needs of the effectiveness of vocational education model reform, and establishes the topological structure of the effectiveness analysis model.

- Input parameters: input parameters have a serious impact on the result output of the model. Different input parameters correspond to different output results. In the effectiveness model of vocational education reform, the input parameters are the duration of the reform, the efficiency of the reform, the teachers of the reform and so on. These factors will affect the output of the effectiveness model. The data collected from the reform of vocational education mode is taken as the input data of the topological structure.
- Output parameter: the output parameter is the specific output result of the effectiveness model under the restriction of different input parameters.

According to the analysis of the input and output parameters of the effectiveness model of vocational education mode reform, the regression function of parameters is established. The response relationship between the input and output functions of the model topology is as follows:

$$y_i = \int_T^i \beta(t) X_i(t) dt + \varepsilon_i \tag{12}$$

where $\beta(t)$ is interval integrable function, $X_i(t)$ is function explanatory variable and ε_i is random error.

For the function response variable $y_i(t)$, it is necessary to establish a regression function to estimate:

$$y_i(t) = X_i^T \beta(t) + \varepsilon_i(t) \tag{13}$$

where $X_i = (x_{i1}, \ldots, x_{ip})$ is the explanatory variable.

In the linear function, if the response variable is of function type, the parameter regression function is:

$$y_i(s) = a(s) + \int_T^i \gamma(s,t)x_i(t)dt + \varepsilon_i(s) \tag{14}$$

where γ is the integrable function.

After building the regression function of input and output parameters of the model, the data mining method is used to mine the indicator data of vocational education reform and complete the model construction. The effectiveness indicator data is too large, the ant colony algorithm, clustering algorithm and random forest algorithm used in the existing model have been unable to meet the needs of the effectiveness analysis of vocational education mode reform (Chang and Meyerhoefer, 2016) so the data mining method is introduced. The semi supervised learning algorithm in the data mining method is mainly used to predict the input data through training, and its performance is mainly determined by the kernel function. At present, the widely used kernel functions are polynomial kernel function, radial basis kernel function and Sigmoid kernel function, whose expressions are

$$\begin{cases} K(x, x_i) = (x * x_i + 1)^p \\ K(x, x_i) = \exp\dfrac{\|x - x_i\|^2}{\delta^2} \\ K(x, x_i) = \tanh\left(\eta\langle x, x_i\rangle + \theta\right) \end{cases} \tag{15}$$

where δ, η and θ are the parameters of kernel function.

Based on the semi supervised learning algorithm introduced above, the model parameters are determined to realise the construction of the effectiveness analysis model of vocational education mode reform (Martimianakis and Hafferty, 2016; Kwan et al., 2015). The specific process is as follows.

The role of semi supervised learning algorithm is to extract the input parameter features of topological structure. The richer the characteristics are, the higher the accuracy of the effectiveness model is. If there is a total of l parameters, the characteristic $x_j^l \ (j = 1, \ldots, N^l)$ of the parameter can be calculated by formula (16):

$$x_j^l = f\left(\sum_{i=1}^N G_{i,j}^l \left(k_{i,j}^l \otimes x_j^{l-1}\right) + b_j^l\right) \tag{16}$$

where $k_{i,j}^l$ and b_j^l represent convolution kernel and feature offset of semi supervised learning respectively, $G_{i,j}^l$ represents feature connection matrix, and function $f(x)$ represents activation function of semi supervised learning.

If l parameters are the feature non offset parameter, and the previous parameter is also a non offset parameter, the calculation formula of the feature vector of the l^{th} parameter is:

$$x^l = f\left(\omega^l x^{l-1} + b^l\right) \tag{17}$$

where ω^l represents the weight of the feature, and b^l represents the offset of the feature.

The semi supervised learning algorithm is used to transform l parameter features into a feature graph. The number of feature graphs is equal to the number of parameter features. Then the calculation method of each feature graph is as follows:

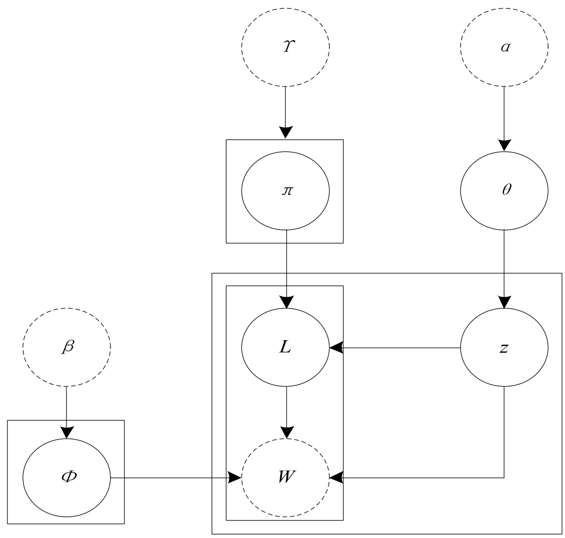
$$x_i^l = f\left(\sum_{i=1}^{N^{l-1}} k_{i,j}^l \otimes x_{i,j}^l + b_j^l\right) \tag{18}$$

Where, the convolution kernel $k_{i,j}^l$ and the characteristic graph $x_{i,j}^l$ are the same size.

After the calculation of feature map, the parameters of effectiveness model need to be determined.

The network diagram of the effectiveness model of vocational education mode reform is shown in Figure 4.

Figure 4 Diagram of effectiveness model of vocational education mode reform network



The parameters of the effectiveness analysis model for the reform of vocational education mode are shown in Table 2.

Table 2 Parameters of effectiveness model of vocational education mode reform

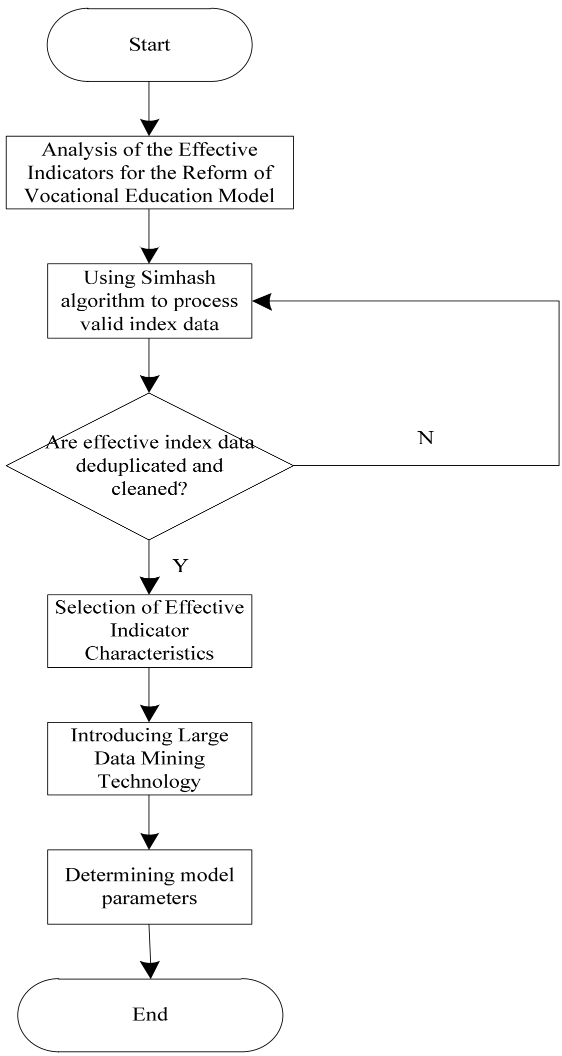
Parameter name	*Parameter weight*	*Parameter description*
n	0.29	Effectiveness indicator data quantity
γ	0.24	Prior parameters
π	0.23	Valid indicator data distribution parameters
I	0.24	Data distribution coefficient

According to the Table 2, it is known that the parameters of the effectiveness analysis model of vocational education mode reform are mainly n, δ, π and I. These four parameters are the basis of model construction, and their values are all vectors. The parameter determination formula is

$$\begin{cases} n = \dfrac{IG(T)}{\Xi} \\ \gamma = \displaystyle\iint \dfrac{H\left(\dfrac{c}{Ri}\right)}{n} \\ \pi = \sum_n \gamma^3 * \xi \\ I = \displaystyle\iiint \pi^2 \pm \gamma \end{cases} \tag{19}$$

where Ξ represents the acquisition factor of effectiveness indicator (Kalaian and Kasim, 2017).

Figure 5 Construction process of effectiveness model of vocational education mode reform



The model can be obtained by substituting the determined parameters into the effectiveness analysis model of vocational education mode reform, and its expression is

$$\xi = \iint \frac{\sum_{i=1}^{n} \gamma * IG(T)}{\pi^{I}} \tag{20}$$

where ξ is the validity indicator.

According to the obtained validity indicator, the output of the model is calculated, and Softmax regression is used to generate the prediction quantity $y = (y_1, \ldots, y_M)^T$ of effectiveness analysis. Where M represents the number of categories, the output result of the effectiveness model is:

$$Y = \frac{e^{-w_i^{L} x^{L-1}}}{\sum_{j=1}^{M} e^{-w_j^{L} x^{L-1}}} \tag{21}$$

where w_i^L is the weight of model Softmax regression.

Then the construction process of the effectiveness analysis model of vocational education mode reform is shown in Figure 5.

Through the above process, the effectiveness analysis model of vocational education mode reform is built (Fayanju et al., 2017).

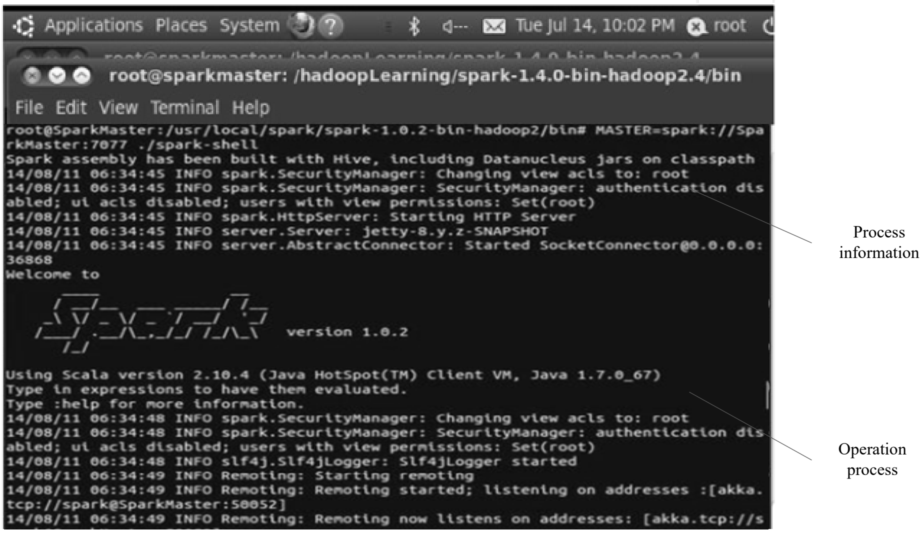
3 Model validation

The above process can realise the construction of the effectiveness model of vocational education mode reform, but it is uncertain whether it can solve the existing problems of the existing model, so the simulation contrast experiment is designed. The experimental data is the teaching mode reform data of 60 vocational education colleges in a province since 2015. In order to increase the accuracy of the experiment, the amount of experimental data selected is twice the amount of data established by the model. In the process of the experiment, the accuracy and effectiveness indicators of effectiveness model analysis are taken as the comparative indicators. The models in Chen and Wang (2018), Xue and Sha (2017), Hu et al. (2018) and the effectiveness model of vocational education mode reform based on data mining are used to carry out the comparative experiment. The specific process is as follows.

3.1 Establishment of experimental environment

The experimental environment is the basis of simulation and contrast experiment. Therefore, the experimental environment is built. This experiment is mainly based on the data mining platform. Two servers are selected to build a Spark cluster, one is recorded as Master and the other as Slave. In the process of building a Spark cluster, because it needs to rely on the file storage system, the default Hadoop framework is built. Based on this, a Spark cluster is built. In addition, in the process of configuring Master and Slave, it needs to keep the file path consistent. The running process of Spark cluster program is shown in Figure 6.

Figure 6 The running process of Spark cluster program (see online version for colours)



Through the above process, the experimental environment is built to provide support for the following experiments.

3.2 Experiment setting

In order to ensure the accuracy of the experimental data, the simulation comparison experiment parameters are set, in which, according to the requirements of the experiment, the data amount of the effective index obtained frequency effective index is set accordingly, as shown in Table 3.

Table 3 Parameter setting

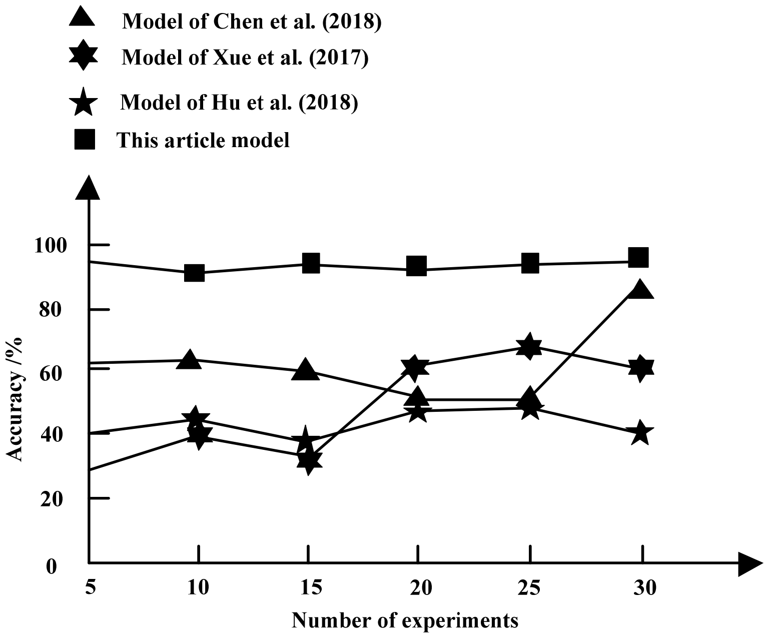
Parameter name	*Parameter significance*	*Parameter setting*
Acquisition frequency of effectiveness indicators	Sampling rate	1.76 Hz
Effectiveness indicator display mode	-	Computer display
Effectiveness indicator data quantity	Acquisition of effectiveness indicators	2.65 MB
Simhash algorithm bits	Execution parameter	32

3.3 Comparison analysis of the accuracy of the model

The accuracy of effectiveness model is an important indicator to verify the overall performance of the model. The analysis accuracy results of the four models are shown in Figure 7.

According to Figure 7, with the increase of the number of experiments, the analysis accuracy of different models will change. When the number of experiments is ten, the analysis accuracy of this model is about 96%, model of Chen et al. (2018) is about 61%, model of Xue et al. (2017) is about 40%; model of Hu et al. (2018) is about 41%; when the number of experiments is 30, the analysis accuracy of this model is about 97.2%, model of Chen et al. (2018) is about 92 %The analysis accuracy of model of Xue et al. (2017) is about 60%; the analysis accuracy of model of Hu et al. (2018) is about 39%; through comparison, we can see that the analysis accuracy of this model is higher and the effect is better.

Figure 7 Analysis accuracy of effectiveness model



3.4 Comparison of the effectiveness indicator of the model

According to the above built experimental environment and experimental parameters, experiments are carried out, and the comparison of effectiveness indicator is shown in Table 4.

As shown in Table 4, the effectiveness index of the model is much higher than the existing three models. It is found that the greater the effectiveness index, the better the model analysis effect. When the number of experiments is 50, the effectiveness analysis model based on ant colony algorithms is 11.59, the effectiveness analysis model based on clustering algorithm is 18.46, the effectiveness analysis model of the efficiency of random forest based algorithms is 12.46, and the effectiveness analysis model based on large data mining is 12.46 The validity index of technology is 22.20; when the number of experiments is 100, the validity analysis model based on ant colony algorithms is 12.89, the validity analysis model based on clustering algorithm is 16.03, the validity of an analysis model of the efficiency of random forest based algorithms is 14.02, and the validity is 14.02 The validity index of the analysis model based on large data mining technology is 26.79; from the experimental results, we can see that the validity index of the model in this paper is higher and the effect is better.

Table 4 Comparison of effectiveness indicator

Number of experiments	*Effectiveness analysis model based on ant colony algorithms*	*Validity analysis model based on clustering algorithm*	*An analysis model of the efficiency of random forest-based algorithms*	*Effectiveness analysis model based on large data mining technology*
10	10.23	11.32	15.23	20.16
20	12.01	13.26	12.00	20.31
30	11.30	12.64	12.06	20.65
40	11.94	16.65	12.05	21.49
50	11.59	18.46	12.46	22.20
60	12.34	17.46	12.23	23.14
70	12.22	17.00	13.33	23.79
80	12.00	16.23	13.59	24.59
90	14.62	16.54	13.49	25.09
100	12.89	16.03	14.02	26.79

4 Conclusions

In order to improve the analysis accuracy of the effectiveness model of vocational education mode reform, this paper constructs the effectiveness model of vocational education mode reform based on data mining, and proves the following conclusions from both theoretical and experimental aspects. When analysing the effectiveness of vocational education mode reform, the proposed model has high analysis accuracy and effectiveness indicators. Specifically, compared with the effectiveness model based on random forest algorithm, the accuracy of the proposed model is greatly improved, and the highest analysis accuracy can reach 97.2%; compared with the effectiveness model based on clustering algorithm, the effectiveness indicator is significantly improved. Therefore, it fully shows that the effectiveness analysis model of vocational education mode reform based on data mining has better analysis effect, and provides model support for the development of vocational education mode reform. However, there is still a large space for the efficiency indicator to rise, which needs further research and optimisation of the model.

References

Adkins, S.J., Rock, R.K. and Morris, J.J. (2017) 'Interdisciplinary STEM education reform: dishing out art in a microbiology laboratory', *Fems Microbiology Letters*, Vol. 365, No. 1, pp.215–227.

Ashby, M.T. and Maher, M.A. (2018) 'The mantra of graduate education reform: why the prayers aren't answered', *Journal of Chemical Education*, Vol. 95, No. 7, pp.1083–1085.

Bini, R.R., Jacques, T.C. and Carpes, F.P. (2017) 'Effectiveness of pedalling retraining in reducing bilateral pedal force asymmetries', *J. Sports Sci.*, Vol. 35, No. 14, pp.1336–1341.

Chang, H.H. and Meyerhoefer, C.D. (2016) 'The causal effect of education on farm-related disability: evidence from a compulsory schooling reform in Taiwan', *American Journal of Agricultural Economics*, Vol. 98, No. 3, p.60.

Chen, J.J. and Wang, J.F. (2018) 'Method for solving Choquet integral model based on ant colony algorithm', *Journal of Shandong University (Engineering Science),* Vol. 48, No. 3, pp.81–87.

Chong, J.L. and Yong, K. (2016) 'Reflection on the education policy orientation in post-May 31 reform in Korea', *Asia Pacific Education Review*, Vol. 17, No. 3, pp.413–426.

Fayanju, O.M., Aggarwal. R., Baucom, R.B. et al. (2017) 'Surgical education and health care reform: defining the role and value of trainees in an evolving medical landscape', *Annals of Surgery*, Vol. 265, No. 3, pp.459–460.

Finge, L.K. (2017) 'Vested interests and the diffusion of education reform across the states', *Policy Studies Journal*, Vol. 46, No. 3, pp.566–578.

Gambarotto, A. and Oken, L. (2017) '(1779–1851): Naturphilosophie and the reform of natural history', *Br. J. Hist. Sci*., 2017, Vol. 50, No. 2, pp.329–340.

Gori, D., Guaraldi, F., Cinocca, S. and Mariapia, F. (2017) 'Effectiveness of educational and lifestyle interventions to prevent pediatric obesity: systematic review and meta-analyses of randomized and non-randomized controlled trials', *Obes. Sci. Pract.*, Vol. 3, No. 3, pp.235–248.

Helma, C., Cramer, T., Kramer, S. and De Raedt, L. (2018) 'Data mining and machine learning techniques for the identification of mutagenicity inducing substructures and structure – activity relationships of noncongeneric compounds', *J. Chem. Inf. Comput. Sci.*, Vol. 35, No. 39, pp.1402–1411.

Hu, J,. Li, S.B., Yu, L.Y. and Yang, G.C. (2018) 'Patent text classification model based on convolutional neural network and stochastic forest algorithms', *Science Technology and Engineering*, Vol. 18, No. 6, pp.268–272.

Kalaian, S.A. and Kasim, R.M. (2017) 'Effectiveness of various innovative learning methods in health science classrooms: a meta-analysis', *Advances in Health Sciences Education*, Vol. 22, No. 5, pp.1–17.

Kim, J. (2018) 'School accountability and standard-based education reform: the recall of social efficiency movement and scientific management', *International Journal of Educational Development*, Vol. 60, No. 21, pp.80–87.

Kwan, J.Y., Nyhof-Young, J., Catton, P., et al. (2015) Mapping the Future: Towards Oncology Curriculum Reform in Undergraduate Medical Education at a Canadian Medical School', *International Journal of Radiation Oncology Biology Physics*, Vol. 91, No. 3, pp.669–677.

Martimianakis, M.A. and Hafferty, F.W. (2016) 'Exploring the interstitial space between the ideal and the practised: humanism and the hidden curriculum of system reform', *Medical Education*, Vol. 50, No. 3, pp.278–280.

Popovic, J.R. (2017) 'Distributed data networks: a blueprint for Big Data sharing and healthcare analytics', *Ann. N.Y. Acad. Sci*., Vol. 1387, No. 1, pp.105–111.

Teixeira, P. and Koryakina, T. (2016) 'Political instability, austerity and wishful thinking: analysing stakeholders' perceptions of higher education's funding reforms in Portugal', *European Journal of Education*, Vol. 51, No. 1, pp.126–139.

Thibault, G.E. (2015) 'Public health education reform in the context of health professions education reform', *American Journal of Public Health*, Vol. 105, No. 1, pp.8–10.

Xue, Y. and Sha, X.Y. (2017) 'On gray prediction model based on an improved FCM algorithm', *Statistics and Decision*, Vol. 2017, No. 9, pp.27–30.