Proceedings of Business and Economic Studies, 2022, Volume 5, Issue 3
http://ojs.bbwpublisher.com/index.php/PBES
ISSN Online: 2209-265X
ISSN Print: 2209-2641

Credit Strategy Design of Small and Medium-Sized Enterprises

Liyi Wang*

Department of Finance, Zhejiang University of Finance and Economics, Hangzhou 310018, Zhejiang Province, China

Corresponding author: Liyi Wang, 992752386@qq.com

Abstract: Small and medium-sized enterprises play an important role in promoting social and economic growth, which is an important foundation for the development of China's national economy. From the perspective of bank interests, this paper uses the invoice data of small and medium-sized enterprises to evaluate their credit risk and formulates corresponding bank credit strategies for these enterprises in different industries and properties. Firstly, the credit risk identification factor system is constructed by using feature engineering, and the quantitative model of enterprise credit risk is built based on back propagation (BP) neural network to predict the default probability of enterprises. On this basis, the k-prototypes clustering algorithm is used to classify enterprises. According to the default probability of each enterprise and the loss rate under different interest rates, a nonlinear programming model, with the maximum expected profit of banks as the goal, is constructed. The simulated annealing algorithm is used to obtain the optimal solution of the credit strategy.

Keywords: Credit strategy; Feature engineering; BP neural network; Simulated annealing algorithm

Online publication: June 20, 2022

1. Introduction

In recent years, with the increasing consumption level, the demand for credit market has gradually increased, along with the proportion of enterprises applying for loans. In order to maintain the normal flow of funds and the normal operation of the system, enterprises frequently seek bank credit. Among them, small and medium-sized enterprises play an important role in promoting social and economic growth, which is an important foundation for the development of China's national economy. Small, medium, and micro enterprises are enterprises with a wide range of social needs, strong employability, and small production scale [1]. However, due to the relatively small scale and insufficient mortgage assets, the main financing channel is to obtain loans from banks through credit guarantee. At the same time, their financial system is flawed, and their credit level is low, thus having weak capabilities to cope with risks. Therefore, when banks make loans to small and medium-sized enterprises, they will evaluate and analyze their credit risk based on their strength and reputation, so as to determine the credit strategy. For banks, loan business is an important part of their asset business. Therefore, it is of practical significance to formulate corresponding credit strategies for small and medium-sized enterprises in different categories.

2. Literature review

Scholars have used a variety of methods to investigate on credit strategy, and certain results have been achieved. In 2008, Fan analyzed the current state of commercial banks' financing for small enterprises and affirmed that increasing credit investment in small enterprises is a strategic method for commercial banks to improve comprehensive income, market competitiveness, and achieve stable as well as sustainable development [2].

In 2010, Chen analyzed the problems of small and medium-sized enterprises' credit, such as the difficulty of credit pricing and the single way of credit guarantee, as well as put forward corresponding suggestions to innovate the credit marketing mode and enrich the credit mortgage guarantee means [3]. In the same year, based on the analysis of the current situation of SMEs and commercial banks in Shaanxi Province, Sun used the method of game theory to investigate the credit problems of commercial banks and SMEs. Information asymmetry, the scale of small and medium-sized enterprises, the probability of successful investment projects, and other factors were taken into account in the model, so as to put forward suggestions for commercial banks to provide credit services to small and medium-sized enterprises [4].

In 2014, Zhao focused on real estate enterprises and analyzed the operation of banks as well as that of real estate enterprises. With the help of an evolutionary game model, through stability analysis and subject strategy selection analysis, he found the reasons for corporate financing constraints, established an interactive selection model, and obtained an evolutionary stability strategy for banks and real estate enterprises, so as to realize the stable development of banks and real estate enterprises [5]. In the same year, from the perspective of credit behavior, Ma expounded the source of credit risk of commercial banks and divided it into a two-participant credit behavior model and three-participant credit behavior model. He used the method of game theory to analyze the behavior strategy of commercial bank credit in an in-depth manner. The game model, in a certain theoretical basis, can be used to guide commercial banks in the credit market to make optimal strategies, avoid credit risk, and ultimately achieve the purpose of stable operation of the credit market [6].

In 2020, Feng studied the lending process of banks to small and medium-sized enterprises, and used the logistics regression model and Bayesian neural network fitting model to establish an optimal credit strategy; that is, to obtain maximum income with minimum risk [7]. In the same year, Zhang established a TOPSIS-RSR combination model from the perspective of banks, quantitatively analyzed the credit risk of a small and medium-sized enterprise, as well as offered the corresponding credit rating results. Following that, a RAROC credit allocation model was established. Fixing the annual total credit, the optimal credit allocation strategy was established based on the credit rating results. Then, a risk factor evaluation model was established to quantitatively evaluate the impact of sudden factors on bank credit risk and analyze specific factors [8].

In 2021, Wang applied the WOE-logistic scoring card model to evaluate the credit risk of small and medium-sized enterprises and suggested an optimal loan strategy [9]. In the same year, Sun used the limited information that banks can easily obtain, such as enterprise transaction information, credit rating, and credit records, and then calculated the expected default rate of small, medium, and micro enterprises under different credit levels by constructing a default rate calculation model of logistic regression. On this basis, a nonlinear programming model of bank optimal credit strategy was constructed, and the model can be used to guide banks to carry out credit activities for potential small and medium-sized enterprise customers [10]. In 2020, Huang studied a bank's credit strategy for small and medium-sized enterprises and established a BP neural network model as well as a grey correlation analysis model with the goal of maximizing bank income expectation [11]. In the same year, Xu constructed a credit risk evaluation index system for small and medium-sized enterprises according to the data of a bank's existing credit record of enterprises and used the projection pursuit model to quantify the credit risk of enterprises. According to Markowitz's portfolio theory, he studied the optimal credit strategy for small and medium-sized enterprises under the principle of risk minimization and income maximization [12].

3. Empirical analysis

3.1. Data source

The data on 123 credit-recording companies and 302 non-credit-recording companies, as well as the statistical data on the relationship between loan interest rates and customer churn rates, including the names of buyers and sellers, as well as the date, status, and tariffs of invoice opening from 2017 to 2019 are all derived from Wind database.

3.2. Factor extraction

The enterprise's sales invoice reflects the original data, thus requiring processing to extract useful information. The extracted factors are shown in **Table 1**.

Table 1. Extracted factors

Factor type	Factor name
Operating cost	growth rate, profit growth rate, sales growth rate, annual average cost growth rate, annual average sales growth rate and annual average profit growth rate
Enterprise credit	return invoice, invalid invoice, valid invoice, return rate, invalid rate, sales invoice, invalid rate
Enterprise scale	invoices, sales invoices, cost growth, income growth
Enterprise information	industry, enterprise nature

According to the name of the enterprise, they are divided into 11 types: individual business, firm, cooperative, head office, limited company, limited company branch, limited liability company, welfare institute, business department, chain limited company branch, and branch. For industries, according to the classification in the Statistical Yearbook of the National Bureau of Statistics, they are divided into nine industries: transportation, warehousing, and postal services; information transmission, software, and information technology services; other services; agriculture, forestry, animal husbandry, and fishery; manufacturing; construction; real estate; wholesale and retail; leasing and business services.

3.3. Data cleaning

Through data exploration, there are some missing values, outliers, inconsistent values, and special symbols in the extracted data. Therefore, data cleaning is done before the actual modeling. The missing data are first interpolated by Lagrange interpolation; the outliers are then processed using the absolute median method. Finally, the data are standardized, and the min-max standardization method is used to process the data to eliminate the influence of dimension.

3.4. Feature engineering

The credit risk of banks is the result of multiple factors; hence, the screening of candidate factors in the construction of classification models is extremely important. The more candidate factors there are, the more comprehensive the analysis of enterprises will be, and the effect of the classification model will also be better. Therefore, sufficient dimensions must be considered in the selection of candidate factors, along with reasonable explanatory. This study mainly used filtering feature selection for factor screening.

Firstly, collinearity characteristics are searched based on the specified correlation coefficient value (setting the threshold value as 0.98), and two collinearity characteristics are removed.

Then, the tree-based machine learning model is used to obtain the feature importance and remove those features with low importance. The label "whether default" is assumed as the dependent variable. The LightGBM library gradient hoist in Python is used to obtain feature importance. A threshold of 0.95 is set to look for the least important features, in which even without these features, a 95% importance can be reached. In order to reduce the variance, the obtained feature importance is the average of 10 rounds of GBM training.

Figure 1 shows the cumulative importance of the corresponding feature number, in which the blue vertical line marks a threshold of 95%. According to the gradient hoist in **Figure 1**, many characteristics cannot contribute to the model. Five low importance factors are identified; the remaining 16 factors can reach a 95% importance.

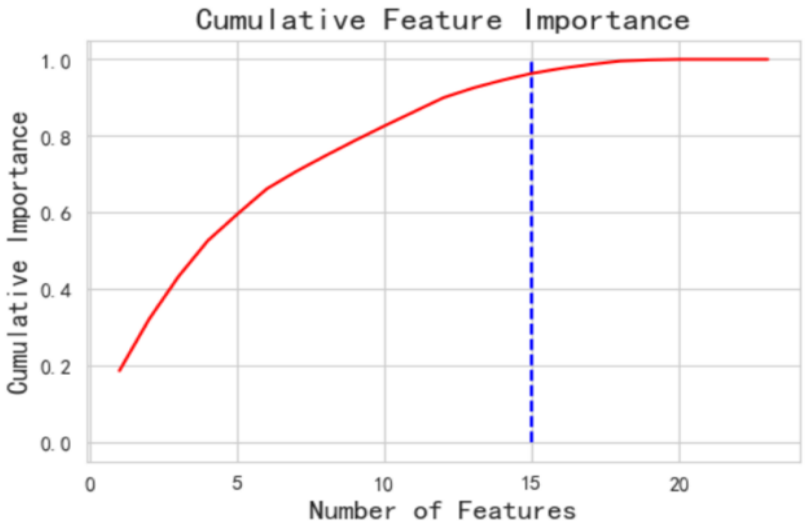

Figure 1. Cumulative importance graph of feature number

After the above steps, 16 factors remained after screening, as shown in **Table 2**. Finally, the modeling data of 123 enterprises (with labels) and 302 enterprises (without labels) are obtained.

Table 2. Factor system

Factor type	Factor name
Operating cost	growth rate, sales growth rate, profit growth rate, average profit growth rate
Enterprise credit	return invoice, return rate, invalid rate, return sales invoice, cancellation invoices, valid sales invoices, return on sales, cancellation rate
Enterprise scale	income growth

3.5. Credit risk identification of enterprises

Since the credit risk of enterprises can be quantified as the repayment probability of enterprises, considering deep learning, BP neural network is constructed and compared with traditional machine learning methods, such as random forest, and high precision is taken as the final model.

In order to ensure that the BP neural network has better training effect, this study adopted the following methods:

(1) Randomly divide the data into a training set and a verification set according to a 7:3 ratio. The random data selection method gives diversity to the BP neural network data.

(2) Each iteration takes n samples, disrupting the order; that is, assuming that there are 80 data, and 10 is taken each time (n = 10), there will be a total of 8 times for all data to be taken.

Since the goal is to predict whether an enterprise defaults or not as well as its probability of default, a more appropriate performance index cross entropy function is used.

$$H(o, y) = -(y \ln o + (1-y)\ln(1-o))$$

Compared with the mean square error, the error of the last layer weight is only related to the difference between the output value and the real value; hence, the convergence will be faster.

The structure of BP neural network is shown in **Table 3**.

Table 3. Structure of BP neural network

Input layer	Latent layer 1 neurons	Latent layer 2 neurons	Output layer neurons	Activation function	Learning rate
16	128	256	2	relu	0.001

Following that, the BP neural network is trained based on Pytorch framework in Python.

Figure 2 shows the curve of the loss value of the changing test set with the number of iterations. It can be seen that the loss value of the test set dropped to about 0.2 after 1,000 iterations. The accuracy, precision, recall rate, F1 value, and AUC of the model are all outputs, and the BP neural network is compared with other models.

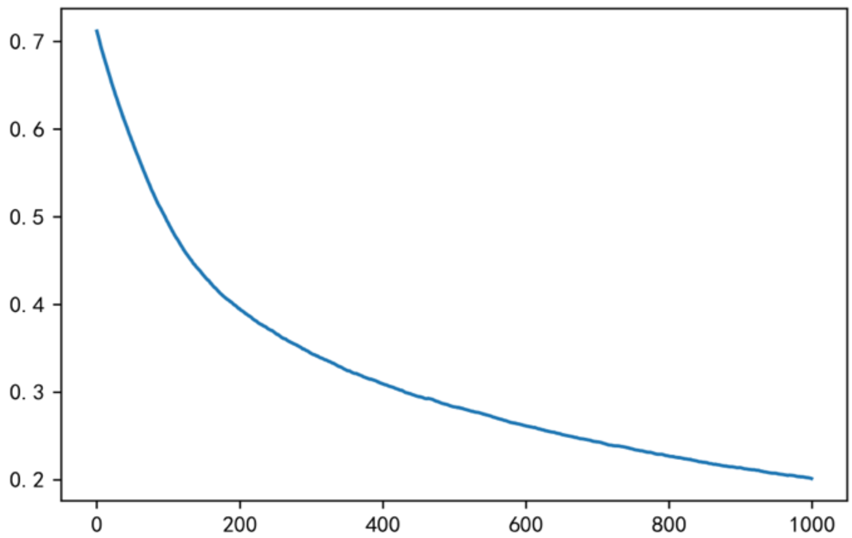

Figure 2. BP neural network iteration diagram

The accuracy of the model is shown in the **Table 4.** The accuracy of the trained model on the test set is 93%, higher than that of the random forest (91%). After synthesizing each index, it is found that the BP neural network is in the optimal state. Therefore, it can be considered that BP neural network takes into account the risks while achieving the maximum excess returns. Therefore, the BP neural network is used as the final model, and the default probability of 123 enterprises is obtained.

Table 4. Model accuracy comparison

Classifier	Accuracy	Precision	Recall rate	F1 value	AUC
Naive Bayes	0.8103	0.8275	0.8000	0.8135	0.8107
Logical regression	0.7758	0.7741	0.8000	0.7868	0.7750
Decision tree	0.8103	0.7878	0.8666	0.8253	0.8083
Random forest	0.9137	0.9310	0.9000	0.9152	0.9142
Bagging	0.8448	0.8620	0.8333	0.8474	0.8452
BP neural network	0.9310	0.9687	0.9117	0.9393	0.9350

3.6. Enterprise grouping

There are many clustering methods. The most common k-means can only deal with numerical data, while k-modes only deal with categorical variable data. However, the data in this study contain both numerical data (such as profit growth rate) and categorical data (reputation rating). In order to simultaneously process two different types of data, this study uses k-prototypes clustering.

In the process of clustering, the k-prototypes algorithm splits the numerical variables and categorical variables as well as calculates the distance between the variables of the samples separately; one uses the Euclidean square distance, while the other uses the Hamming distance (the number of corresponding bits between two variables with the same number of bits). Both are added as the distance between samples.

The determination of k value is crucial for k-prototypes clustering. However, since the number of clusters determined by experience is too large and it is not necessarily the real number of clusters, we hope to determine the real number of clusters from the data itself; that is, the optimal number of clusters for data. In this paper, the loss function F of k-prototypes is used as the evaluation index to calculate the error of k from 2 to 10.

$$F = \sum_{l=1}^{k} \sum_{i=1}^{n} w_{li} \cdot D(x_i, z_i)$$

Among them, 1 indicates that the data object x_i is divided into the first class cluster; 0 indicates that the data object x_i is not divided into the first class cluster, but it is the difference of numerical properties; F is the clustering error of all samples, representing the clustering effect.

The core idea is that as the number of cluster k increases, the sample division will be more refined, and the degree of aggregation of each cluster will gradually increase, so the loss function F will naturally become smaller. Additionally, when k is less than the real number of clusters, the increase of k will greatly increase the degree of aggregation of each cluster, so the decline of F will be large, and when k reaches the real number of clusters, the return of the degree of aggregation obtained by increasing k will decrease rapidly, so F will decrease sharply, and then flatten as the k value continues to increase; hence, there is an inflection point in the relationship graph between F and k, in which the k value that corresponds to this inflection point is the real number of clusters of data.

With Python, let k be from 2 to 10, cluster each k value, note the corresponding F, and then draw the relationship between k and F.

It can be seen from **Figure 3** that the decline of F begins to slow down when k = 4, where this point is the inflection point. Therefore, corresponding to the inflection point, 4 is selected as the best clustering number.

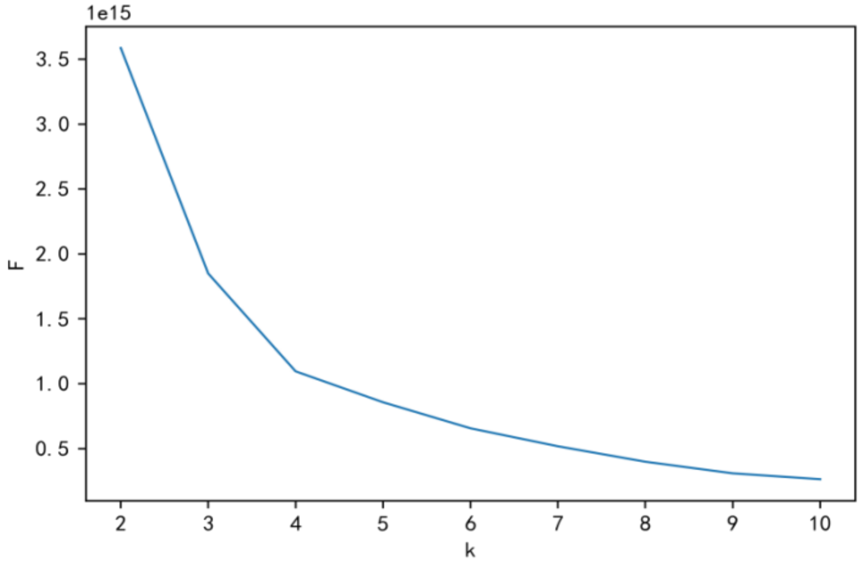

Figure 3. k value diagram

Using Python, call the function in kmodes.kprototypes library for k-prototypes clustering, set k = 4, obtain the class label of each enterprise, and output the clustering center, as shown in **Table 5**.

Table 5. Enterprise cluster center table

Factor	Category 1	Category 2	Category 3	Category 4
return invoice	6	8	2	7
import return rate	0.0051	0.0095	0.0039	0.0078
…	...	...	...	...
annual average profit growth rate	0.4255	-0.1457	0.0725	0.3069
credit rating	A	C	B	B

According to **Table 5**, the enterprises are classified into high-end customers, potential customers, mid-end customers, and low-end customers.

High-end customers (category 1) are customers who have the characteristics of high effective sales invoices, annual average profit rate, and high income growth rate. This shows that the enterprise yield of this kind of customer is high; their assets have high liquidity, their capability to repay loans is strong, their effective invoice is relatively large, and they have good reputation.

Potential customers (category 3) are customers who have high average annual profit rate and income growth rate, but their effective sales invoice volume is relatively lower than that of high-end customers. This enterprise return rate of this kind of customer is higher than the enterprise average level; their repayment ability is strong, but their effective invoice quantity is relatively small.

Mid-end customers (category 4) are customers who have positive annual profit growth rate and income growth rate, but they have higher invoice rejection rate, lower enterprise credit, and lower possibility of timely repayment of credit.

Low-end customers (category 2) are customers who have negative average annual profit rate and income growth rate; their return rate of sales is higher, but their effective sales invoice is also higher. This reflects that although low-end customers have good corporate reputation, their capital base is weak, their profitability is weak, and the possibility of overdue repayment of credit is high. Excessive loans to them will cause losses to banks.

3.7. Construction of a bank profit model

A bank's profit comes from the additional interest after lending to an enterprise, but when the enterprise does not repay, the loan will be regarded as a bad debt, and not as a loss. According to the above model, the probability *t* of enterprise repayment can be obtained, so in the case of only considering whether the enterprise will repay, it is then possible to obtain the expectation of profit *p* obtained by the bank from enterprise *i* when the interest rate is *lr* and the loan amount is *z*.

$$E(p) = lr_i \cdot z_i \cdot t_i - (1-t_i) \cdot z_i$$

In addition, the probability $w = r$ that an enterprise may choose other banks at a fixed interest rate can be obtained from the customer churn rate *r* of enterprises with different credit levels. When an enterprise chooses not to borrow from the bank, the profit that the bank can obtain from the enterprise is 0; however, when the enterprise chooses to borrow from the bank, the profit obtained by the bank from the enterprise is $lr_i \cdot z_i$. Therefore, considering the probability of customer churn and repayment, the expectation of the profit obtained by the bank from an enterprise is as follows:

$$E(p) = w_i \cdot (lr_i \cdot z_i \cdot t_i - (1-t_i) \cdot z_i)$$

In the previous model, a clustering method for data is provided. In order to facilitate the management of banks, the same borrowing strategy can be used for enterprises under the same category. Then, the total profit of banks can be regarded as the sum of profits obtained from each category. The constraint condition is that the total loan amount is not higher than the given total amount, and the loan amount of each enterprise is between 10 and 1 million, while the interest rate is between 0.04 and 0.15. The optimization model of this strategy is as follows:

$$E(p) = \sum_{i=1}^{m} \sum_{j=1}^{n} w_{ij} \cdot (lr_{ij} \cdot z_{ij} \cdot t_{ij} - (1-t_i) \cdot z_i)$$

The constraint condition is as follows:

$$\begin{aligned} &0.04 \le lr_i \le 0.15 \\ &10 \le z_i \le 100 \\ &\max \sum_{i=1}^{m} \sum_{j=1}^{n} z_{ij} \le \text{Total annual bank credit} \end{aligned}$$

3.8. Solving the bank profit model based on simulated annealing algorithm

The optimization model of bank profit is a nonlinear constrained optimization problem. The interest rate that can be obtained by banks is not continuous, so the commonly used gradient optimization methods, Lagrange function, and Hamiltonian function are not available in this optimization model. Therefore, the simulated annealing algorithm is used to solve the optimization problem.

The main process of simulated annealing algorithm in this paper is as follows:

(1) randomly select the initial value *x* in the feasible region, calculate the initial target value *f(x)*, and set the initial value;

(2) produce a perturbation term *dx* and obtain a new solution $x_1 = x + dx$; repeat step 2 if x_1 does not satisfy the constraint condition;

(3) comparing $f(x)$ and $f(x_1)$, based on the metropolis principle, when $f(x_1) > f(x)$, the value of x_1 is given to x, when $f(x_1) < f(x)$, calculate, and randomly generate a number i; if $i \leqslant p$, then use the value of x_1 to replace the original value of x; set the number of turns L, reduce the temperature T after each L; $T = \alpha T$, where the constant α assumes 0.9 as the annealing rate; if the temperature T does not meet the condition $T < T_{end}$, the second step is returned.

In order to ensure that the amount of constraint can play a role, this study sets the total amount of bank loans as 70 million, the number of turns L as 100, the initial temperature T = 100, and termination temperature T = 0.01. The final results are shown in **Table 6**.

Table 6. Optimal strategy

Category	Lending amount (ten thousand RMB)	Enterprise number	Lending rate
High-end customers	84	26	0.04
Mid-end customers	57	41	0.0465
Potential customers	55	27	0.0425
Low-end customers	30	29	0.0465
Total credit (ten thousand RMB)		6876	
Optimal profit (ten thousand RMB)		63.34	

From the optimal strategy for these enterprises, banks should assume several measures.

(1) For high-end customers, there should be a large number of loans and low interest rates to protect the loss caused by non-payment, and as far as possible, banks should retain this type of customers.
(2) The number of mid-end customers is relatively large. These customers can take a low loan amount, high interest rate measures.
(3) The number of potential customers is small, so banks can allocate higher loan amounts and reduce interest rates to retain such customers.
(4) Low-end customers tend to take lower loan amounts; therefore, having high interest rates and small loans may reduce the loss of possible bad debts.

4. Conclusion

Aiming at the credit decision issue of small and medium-sized enterprises, this study constructs a credit risk identification factor system and a quantitative model of enterprise credit risk based on BP neural network to predict the default probability of enterprises. On this basis, the clustering model is used to classify enterprises. According to the default probability of each type of enterprise and the loss rate under different interest rates, a nonlinear programming model, with the maximum expected profit of banks as the goal, is constructed. Simulated annealing algorithm is used to obtain the optimal solution of the credit strategy.

With regard to this model, the following conclusions can be drawn: the accuracy of BP neural network can reach 93%, and it can accurately identify credit risk; k-prototypes clustering is scientific and reasonable for enterprise division; the optimization model based on clustering greatly reduces the convergence speed of the model, so that the bank can be conveniently managed; the simulated annealing algorithm can be used to find the global optimal solution. The final bank credit strategy provides a theoretical basis for banks to develop credit strategies, which has practical significance.

In addition, this optimization process can be extended to general banks to solve optimal credit policy problems and other optimization problems, such as the selection of the shortest path and capacity maximization issues. In addition, combined with the optimization process of clustering in this study, it is possible to solve socio-economic problems, such as the maximization of production capacity of a certain group and the reduction of unnecessary losses in the society, as well as simplify calculation costs.

Disclosure statement

The author declares no conflict of interest.

References

[1] Wang L, Li H, 2019, Construction and Empirical Test of Credit Evaluation System for SMEs in China. Journal of Jilin Institute of Chemical Technology, 36(06): 86–92.

[2] Yuan K, Li Z, Zhao J, 2008, Research on Credit Strategy of Small Enterprises. Hebei Finance, 2008(04): 45.

[3] Chen M, Li J, Zhang L, 2010, The Credit Strategy of SMEs. Modern Finance, 2010(11): 31.

[4] Sun Y, 2010, Research on Credit Strategy of Small and Medium-sized Enterprises in Commercial Banks – An Empirical Analysis of Shaanxi Province. Journal of Xi'an University of Finance and Economics, 23(04): 31–35.

[5] Zhao J, 2014, Research on the Win-Win Credit Strategy of Cooperation Between Banks and Real Estate Enterprises Based on Evolutionary Game Theory. Yunnan Normal University.

[6] Ma Q, 2014, Research on Credit Behavior and Credit Strategy of Commercial Banks in China Based on Game Theory. Northeastern University.

[7] Feng Y, Xiao L, Zheng D, 2020, Research on Credit Strategy of Small and Medium-Sized Enterprises Based on Bi-Objective Programming. Mall Modernization, 2020(20): 111–113.

[8] Zhang Q, Hu J, Gu C, 2020, Research on Credit Strategy Based on TOPSIS-RSR Combination Model. Mall Modernization, 2020(22): 181–183.

[9] Wang Z, Yang X, Wu J, 2021, Research on Credit Strategy of Small and Medium Enterprises. Mathematical Modeling and Application, 10(01): 80–91.

[10] Sun Y, 2021, Research on the Optimal Credit Strategy of Banks to Small and Medium-Sized Enterprises Under Information Asymmetry – Based on Logistic Regression Default Rate Measurement Model. Financial Development Research, 2021(06): 78–84.

[11] Huang X, Zhu R, Qin Y, 2022, Research on Bank Credit Strategies for Small and Medium-Sized Enterprises. China Business Theory, 2022(03): 76–78.

[12] Xu Y, Zhu J, 2022, Research on the Optimal Credit Strategy of SMEs Based on Projection Pursuit Model. Journal of Qingdao Agricultural University (Social Science Edition), 34(01): 47–50.