Research on Credit Decision Issues of the Small and Medium-Sized Enterprises Based on TOPSIS and Hierarchical Cluster Analysis

Zhipeng WANG[a] and Bo LIN[b,1]

[a] *School of Computer and Electronic Information, Guangxi University, China*
[b] *School of Mathematics and Information Science, Guangxi University, China*

Abstract. Small and medium-sized enterprises have promoted the development of the national economy, but they often ask for loaning from banks due to the shortage of funds and mortgage assets. Therefore, it is very important for banks to formulate reasonable credit strategies. This paper introduces TOPSIS and hierarchical cluster analysis algorithm, establishes a credit decision-making model based on the strength, stability of supply and demand, reputation and anti-risk ability of enterprises, quantitatively scores the risk of each enterprise, then obtains the lending rates, and establishes an objective function through RAROC theory to obtain the loan amount when the bank's profit is the maximum. In addition, the missing indicators of enterprises without credit records can be predicted by BP neural network prediction model. This paper studies and solves the credit decision-making problems of small and medium-sized enterprises, and provides a reference for banks to put forward more accurate credit strategies.

Keywords. TOPSIS, Hierarchical cluster analysis, BP neural network, Credit decisions, RAROC theory

1. Introduction

As an important part of the national economy, small and medium-sized enterprises play an indispensable role in the vigorous development of Chinese economy. Since the reform and open, Chinese economy has developed rapidly. During this period, a large number of small and medium-sized enterprises have emerged, including those with good operation and those on the verge of bankruptcy. They have to loan from banks for various reasons. For banks, their funds are limited. Therefore, how to formulate appropriate credit strategies and allocate resources reasonably is the key to the healthy development of Chinese economy. In the direction of quantitative credit scoring and determining credit strategies for small and medium-sized enterprises, a large number of theoretical and practical research have been carried out and rich research results have been accumulated.

---

[1] Corresponding Author, Bo LIN, School of Mathematics and Information Science, Guangxi University, China; Email:2033074197@qq.com.

Xiong Yudi [1] made a quantitative analysis of the scale of the enterprise based on data analysis and analytic hierarchy process, which provided a certain reference for the bank loan strategy. Xing Xudong [2]adopted the decision tree model based on logistic analysis method to quantitatively analyze the enterprise from many aspects such as default probability and invoice information. Compared with the traditional method, the error rate is reduced by 60.2% - 70.1%. Nan Jinxiang, Li Wenzhe [3] and others used TOPSIS method to obtain the comprehensive score of each enterprise by using seven indicators, and considered the customer churn rate to solve the optimal credit decision. Liu Xinying [4] carried out objective optimization on the basis of logistic model, considered the loan amount, lending rates and customer churn rate, and found the optimal interest rate to enable the bank to obtain the maximum investment return. Zhang Bingqing [5] used TOPSIS comprehensive evaluation model to obtain the credit score of each enterprise. Finally, he defined and gave a specific linear strategy model to formulate the credit strategy for small and medium-sized enterprises.

Although the above research uses different credit decision-making models, this paper uses systematic cluster analysis to effectively determine whether to grant loans to enterprises. At the same time, neural network is used to predict the missing indicators of enterprises without credit records. According to RAROC theory, factors such as bank risk cost, capital cost and customer churn rate are considered at the same time. It further optimizes the credit decision-making model and improves the accuracy and effectiveness of the model.

2. Model Construction And Solution

In order to quantitatively analyze the credit risk of several companies, the strength of each enterprise, the stability of supply and demand, the sensitivity of lending rates and credit situation are taken as evaluation indicators. Based on the credit data of these companies, a credit decision-making model is established to formulate accurate loan strategies for small and medium-sized enterprises.

2.1. Data Selection

The data in this paper comes from the relevant data of several enterprises with credit records (credit records contain two indicators of credit grade and whether to default) and no credit records of a bank in China, such as input and output invoice information, including enterprise code, invoice number, invoicing date, input and output unit code, amount, tax amount, total amount and tax, invoice status information, and the relationship between the annual interest rate of bank loans and customer churn rate. The loan amount is required to be between 100000 yuan and 1 million yuan, and the annual interest rate of the loan is required to be between 4% and 15%.

2.2. Data Preprocessing

Refine the indicators according to the data information, that is, extracting the following indicators from the original data information, namely, the total input tax t_1, the total output tax t_2, the total number of input and output invoices t_3, the total effective amount of input invoice t_4, the total effective amount of output invoice t_5, the effective rate of input invoice t_6, the effective rate of output invoice t_7, the ratio of total input and output

price tax t_8, the ratio t_9 of the total number of effective invoices of input and output, the ratio t_{10} of the total amount of effective invoices of input and output, the credit rating t_{11}, whether to default t_{12}, and the enterprise scale t_{13}.

- Quantification of credit rating

The credit ratings of the above companies are divided into four categories: A, B, C and D. if the credit rating is A, $t_{11} = 4$; If the reputation level is B, $t_{11} = 3$; If the credit rating is C, $t_{11} = 2$; Credit rating is 1; Then $t_{11} = 1$.

- Quantification of whether to default

Encode the default indicator. If default, $t_{12} = 0$; If there is no breach, $t_{12} = 1$.

- Convert all indicators to very large indicators

It is found that only the ratio t8 of the total price and tax of input and output, the ratio t9 of the total effective invoices of input and output, and the ratio t10 of the total effective amount which involves input and output invoices are intermediate indicators, and the rest are very large indicators. Therefore, these three indicators are transformed into very large indicators:

Set the intermediate index sequence as $(x_1, x_2, ..., x_n)$

Then the transformed very large index sequence is $(y_1, y_2, ..., y_n)$

$$y_i = 1 - \frac{|x_i - x_{best}|}{\max\{|x_i - x_{best}|\}} (i = 1, 2, 3, ..., n) \tag{1}$$

2.3. Solving Process of Credit Decision Model for Enterprises with Credit Records

Step 1: Construct TOPSIS evaluation function TOPSIS (x). X is the incoming forward matrix. The following is the specific algorithm of TOPSIS evaluation function:

Standardize the forward matrix, M enterprises and N evaluation indicators, so there is $M \times N$ matrix. Due to different dimensions, the matrix must be standardized to eliminate the influence of dimensions.

Set forward matrix:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdot & \cdot & \cdot & x_{1n} \\ x_{21} & x_{22} & \cdot & \cdot & \cdot & x_{2n} \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ x_{m1} & x_{m2} & \cdot & \cdot & \cdot & x_{mn} \end{bmatrix} \tag{2}$$

Then the element in the standardized matrix Y is:

$$y_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{n} x_{ij}^2}} \tag{3}$$

Calculate and normalize the score, abstract two ideal vectors, and abstract an evaluation object with the lowest evaluation index and an evaluation object with the highest evaluation index from all evaluation index data. The following is the normalization process of the standardization matrix:

Standardized matrix and row vector form of its evaluation object:

$$
Y = \begin{bmatrix} y_{11} & y_{12} & \cdot & \cdot & \cdot & y_{1n} \\ y_{21} & y_{22} & \cdot & \cdot & \cdot & y_{2n} \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & & \cdot & \cdot \\ y_{m1} & y_{m2} & \cdot & \cdot & \cdot & y_{mn} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \cdot \\ \cdot \\ \cdot \\ \alpha_m \end{bmatrix} \tag{4}
$$

Define maximum and minimum:

$$
y^{best} = (\max\{y_{11}, y_{21}, ..., y_{m1}\}, ..., \max\{y_{1n}, y_{2n}, ..., y_{mn}\}) \tag{5}
$$

$$
y^{worst} = (\min\{y_{11}, y_{21}, ..., y_{m1}\}, ..., \min\{y_{1n}, y_{2n}, ..., y_{mn}\}) \tag{6}
$$

Get the score of the i-th (i=1,2, ... ,n) evaluation object without normalization:

$$
s_i = \frac{\left| \alpha_i - y^{worst} \right|}{\left| y^{best} - \alpha_i \right| + \left| \alpha_i - y^{worst} \right|} \tag{7}
$$

The scores were normalized:

$$
\tilde{s}_i = \frac{s_i}{\sum\limits_{i=1}^{n} s_i} \tag{8}
$$

That is, the column vector composed of the last score is used as the return value of TOPSIS function:

$$
S = (\tilde{s}_1, \tilde{s}_2, ..., \tilde{s}_m) \tag{9}
$$

Step 2: Select characteristic variables and build a credit risk quantitative model. Then decide whether to grant loans to enterprises according to enterprise strength and the stability of enterprise supply-demand relationship.

- Enterprise strength

The measurement of enterprise-strength depends on a variety of indicators. Here, the total input tax t_1, the total output tax t_2, the total number of input and output invoices t_3, the total effective amount of input invoice t_4, the total effective amount of output invoice t_5 of each enterprise are selected as the measurement standards of enterprise-strength, and the TOPSIS calculation method is used for scoring.

For enterprise j, the enterprise-strength score is:

$$ES_j = \text{TOPSIS}(t_1, t_2, t_3, t_4, t_5) \qquad (10)$$

- Stability of enterprise supply-demand relationship

The stability of the enterprise supply-demand relationship also depends on a variety of indicators. Here, the ratio t_8 of the total price and tax of input and output, the ratio t_9 of the total effective amount of input and output invoices, and the ratio t_{10} of the total effective amount of input and output invoices are selected as the measurement standards of the stability of enterprise supply-demand relationship. TOPSIS algorithm is also used to score the stability of the enterprise supply-demand relationship.

For enterprise j, the stability score of the enterprise supply-demand relationship is:

$$ESD_j = \text{TOPSIS}(t_8, t_9, t_{10}) \qquad (11)$$

- Enterprise anti-risk ability

The anti-risk ability of enterprises of different sizes is different. Therefore, enterprise size t_{13} is selected as the measurement standard of enterprise anti-risk ability, and the TOPSIS algorithm is also used to score the enterprise anti-risk ability.

For enterprise j, the enterprise anti-risk ability is:

$$EAR_j = \text{TOPSIS}(t_{13}) \qquad (12)$$

The risk is quantified according to the four aspects of enterprise-strength, stability of enterprise supply-demand relationship, enterprise reputation, and anti-risk ability. It is known that the score of enterprise credit risk is a very small index, so the score of enterprise-strength, stability of enterprise supply-demand relationship, enterprise reputation, and anti-risk ability should be transformed into a very small index first. Then the enterprise credit risk score obtained according to TOPSIS algorithm is:

$$ECR_j = \text{TOPSIS}(ES_j, ESD_j, ER_j, EAR_j) \qquad (13)$$

Step 3: Firstly, according to the enterprise strength and the stability of the enterprise supply-demand relationship, the TOPSIS algorithm is used to obtain the score of loan or not. Then, according to the obtained loan or not, we make a hierarchical cluster analysis with a classification number of 2, and we decide to make loans to the classes with high scores.

Calculate the score of loan or not:

$$LE_j = \text{TOPSIS}(ES_j, ESD_j) \qquad (14)$$

Now, the 123 groups of indicators are subject to variable clustering, that is R-type clustering analysis. The individual distance adopts square Euclidean Distance, the system clustering method adopts inter-group connection, and the cluster number is set to 2. According to the solution of SPSS software, 95 enterprises are clustered into one category, and the score of loan or not is high, and the other 28 enterprises are clustered into one category, and the score of loan or not is low. Therefore, loans were granted to the 95 enterprises. Because the total clustering pedigree is too large, we intercepted the

cluster dendrogram of 28 enterprises that do not issue loans. The figure 1 is the Cluster Dendrogram of non lending enterprises.
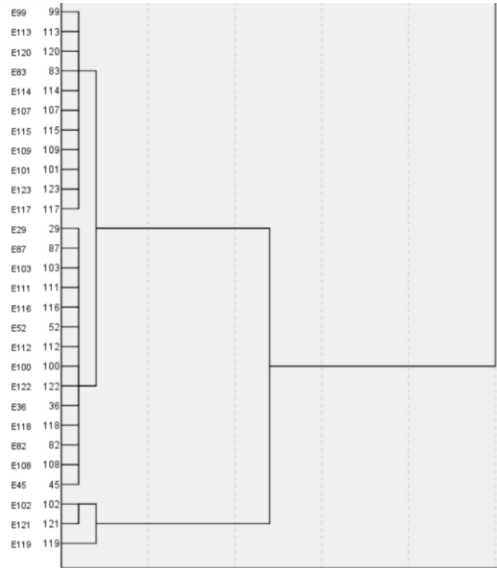

Figure 1. the Cluster Dendrogram of non lending enterprises

For enterprise j, the enterprise lending rates is:

$$Eloan_j = 4\% + \frac{ECR_j - ECR_{j\min}}{ECR_{j\max} - ECR_{j\min}} \bullet (15\% - 4\%) \tag{15}$$

Then we calculate the Lending rates of 95 enterprises according to formula 15. The result is shown in table 1.

Table 1. Lending rates

Enterprise	Lending rates	Enterprise	Lending rates
E1	0.04000	E86	0.14123
E2	0.11432	E87	0.14047
E3	0.12440	E88	0.14015
E4	0.11367	E89	0.14055
E5	0.13508	E90	0.14002
E6	0.13105	E91	0.14954
E7	0.12709	E92	0.14065
E8	0.12198	E93	0.15000
E9	0.13374	E94	0.13986
E10	0.13528	E95	0.14062

Step 5: Determination of loan amount. According to RAROC [6] theory, we establish a function to maximize the bank's profit.

According to RAROC loan pricing theory, this paper considers capital cost and risk cost as the evaluation basis of loan pricing, and also considers the ratio that banks lose potential customers due to factors such as lending rates, that is, customer churn rate Li. Based on the above aspects, this paper constructs the function of Bank net profit.

$$\max f = (1 - L_i) \bullet \sum_{i=1}^{95} (A_i \bullet r_i - A_i \bullet FTP - EL_i) \tag{16}$$

A_i represents the loan amount granted by the bank to the enterprise code i. R_i is the bank's lending rates to the enterprise code i. FTP is the benchmark interest rate, where the benchmark interest rate takes the Shibor to mean value of Bank of Shanghai on November 19, 2021, i.e. 2.756%, EL_i represents the risk cost of the enterprise with enterprise code i.

Here, risk cost (EL_i), also known as anticipated loss, is an expected loss of loan default in advance. It consists of risk exposure A, anticipated default rate EDF, and default loss rate LGD:

$$EL_i = A_i \cdot LGD_i \cdot EDF_i \tag{17}$$

According to relevant economic literature [6], the default loss rate is:

$$LGD_i = \beta \cdot A_i \cdot r_i \tag{18}$$

Among them, β is the default loss coefficient. According to the default pyramid principle [7], the default loss coefficient is a piecewise function of rating, in which the value of class an enterprise can be 0.09, class B enterprise can be 0.18, and class C enterprise can be 0.27.

We represent the average default rate of 123 enterprises as the general default rate, i.e:

$$cl = \frac{1}{123} \sum_{i=1}^{123} t_{12} \tag{19}$$

Then the default rate of 95 enterprises can be obtained according to the credit risk score and general default rate of the selected 95 enterprises, that is, for enterprise j:

$$EDF_i = cl \cdot ECR_i \cdot 123 \tag{20}$$

Generally, the amount of loans granted to small and medium-sized enterprises is between 100000-1 million:

$$10 \leq A_i \leq 100 \tag{21}$$

Using lingo to solve, the specific loan amounts of 95 enterprises are shown in the table 2.

Table 2. Loan Amounts

Enterprise	Loan amount	Enterprise	Loan amount
E1	95.26715	E86	55.40541
E2	87.20867	E87	16.63750
E3	69.09409	E88	53.60359
E4	67.18400	E89	18.69118
E5	79.11071	E90	54.34436
E6	87.35255	E91	51.30493
E7	86.40460	E92	51.30493
E8	86.86789	E93	18.96898
E9	85.72270	E94	52.45277
E10	51.70463	E95	10.05894

2.4. Solving Process of Credit Decision Model for Enterprises without Credit Records

Compared with enterprises with credit records, enterprises without credit records only lack two indicators: reputation level and default. Neural network algorithm can be used to predict the reputation level and default of enterprises without credit records.

Step1: set the input layer to the data of non predicted variables of 123 groups of enterprises with credit records: total input tax and output tax, total number of input and output invoices, total effective amount of input invoice, total effective amount of output invoice, total input price tax, total output price tax, input invoice efficiency and output invoice efficiency. The output layer is 123 groups of credit rating and loan data to be predicted, so as to train the model.

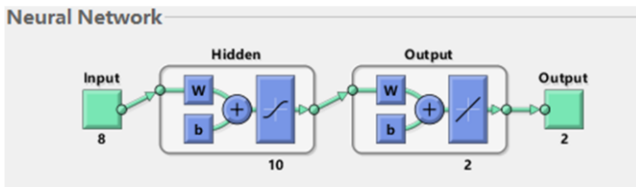The figure 2 shows the neural network diagram of the BP neural network used.

Figure 2. Graph of BP neural networkhort caption

Step 2: MSE diagram is obtained after neural network training with MATLAB. Generally speaking, the more training times, the smaller the error. However, due to the phenomenon of over fitting, the mean square error of the validation data set may increase. Therefore, the minimum value of MSE after a training should be found. The figure 3 shows the MSE diagram in the neural network training process of credit rating and loan.
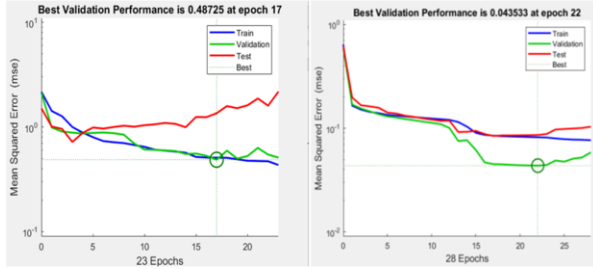
Figure 3. MSE of reputation level and loan or not indicator

The training of neural network is to continuously adjust the weight matrix. When the left and right images are trained 17 and 22 times respectively, they all correspond to the minimum MSE. The MSE of the two figures are 0.48725 and 0.043533 respectively, which are relatively small errors for the original data, which can better reflect the accuracy of the credit model.

Step 3: Error analysis after MATLAB input neural network training

Regression between the fitting value and the real value shows that the R values of both are more than 0.7, that is, the goodness of neural network fitting is better. The figure 4 shows the R values of two indicators.
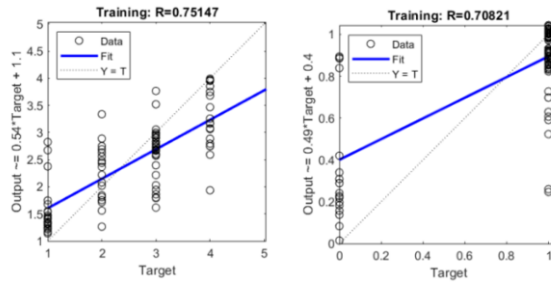

Figure 4. R values of reputation level and loan or not indicator

Save the trained neural network model and results, and then use the SIM function of MATLAB for prediction. After the preliminary results are obtained, all values are rounded, that is, the data of the final missing reputation level and whether to default are obtained.

Step4: after obtaining the two indicators of missing credit rating and default, it is transformed into the solution process of credit decision model of enterprises with credit records.

2.5. The Reasonableness Test of Model

At the heart of the credit decision model established in this paper is the creation of a credit risk quantification model, i.e. the solution to the risk score. To examine the plausibility of the risk score, we make a scatter plot of the relationship of credit risk score and credibility level. The scatter plot is shown in figure 5.
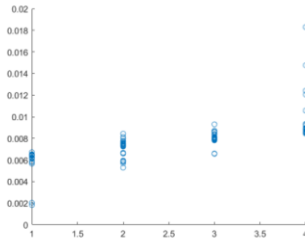
Figure 5. Scatter plot of the relationship of credit risk score and credibility level.

Where the abscissa is the credibility level and the ordinate is the credit risk score. We used the Spearman coefficient r_s to measure the correlation between two variables, i.e., to test whether it was significantly different from 0. Setting the null hypothesis H0: $r_s = 0$ and the alternative hypothesis H1: $r_s \neq 0$, hypothesis testing was performed using SPSS software. The results of hypothesis test are shown in figure 6.

Correlation

			Credit risk score	Credit rating
Spearman Rho	Credit risk score	Correlation coefficient	1.000	.897**
		Sig. (Two-tailed)	.	.000
		N	115	115
	Credit rating	Correlation coefficient	.897**	1.000
		Sig. (Two-tailed)	.000	.
		N	115	115

**. At the 0.01 level (two tailed), correlations were significant.

Figure 6. Hypothesis test resultes

At the 99% confidence level, we can reject the original assumption that H0: rs = 0, that is, rs is significantly not zero.

Therefore, the credit risk score and credit rating have a significant positive correlation, so the establishment of our credit risk quantitative model is effective and accurate, which further proves the accuracy of the credit decision-making model.

3. Conclusion

3.1. Advantages and promotion of the model

Based on the related data of small and medium-sized enterprises in a bank database, this paper establishes a credit risk assessment model by using TOPSIS and system clustering model, solves the problem of setting lending rates, and effectively establishes an optimal loan decision model by using RAROC theory and default pyramid theory. Then the loan amount allocated to enterprises when the bank profit is the largest is solved. In addition, when small and medium-sized enterprises have no credit records, they can also use BP neural network to predict the credit records according to various indicators in the input and output invoice information, so as to better formulate the credit strategy for small and medium-sized enterprises. The model can be applied not only to the credit decision-

making of banks, but also to more scenarios, such as comprehensive evaluation of hospitals in a city, talent recruitment of an enterprise and so on.

3.2. The Future of the Model

In the credit decision model, we can take into account the possible impact of some small and medium-sized businesses, and introduce contingencies to improve the credit decision making model, such as national policies and COVID-19 emergencies. More machine learning algorithms can also be used to improve the model, such as expanding the model from the perspective of game theory. In addition, we can consider more costs of banks, establish a more scientific and more based profit function, and build a more accurate credit decision-making model.

References

[1] Xiong Yudi, Liu Genmao, Lin Zhitong, Wei Baoying. Solving credit decision problems based on data analysis and analytic hierarchy process [J]. China collective economy, 2021 (27): 75-76.
[2] Xing Xudong. Analysis on bank credit strategy for small, medium and micro enterprises [J]. Business news, 2021 (30): 79-81.
[3] Nan Jinxiang, Li Wenzhe, Zhao Zhankui, Liu Chang, Wang Huilin. Research on credit decision of small, medium and micro enterprises based on TOPSIS [J]. Business news, 2021 (15): 119-120.
[4] Liu Xinying, Zhang Ziyi, Liu Yuqing, Li Mengjie, Yao Jingdu. Research on credit decision-making of commercial banks for small, medium and micro enterprises -- Taking Title C of 2020 National College Students' mathematical modeling competition as the background [J]. Modern marketing (Business Edition), 2021 (03): 24-25.
[5] Zhang Bingqing. Determining credit strategy based on TOPSIS Model and logistic classification model [J]. Fortune today, 2021 (11): 31-32.
[6] Gao Jiaheng. Research on loan pricing of small enterprises based on RAROC [D]. Zhejiang University, 2013.
[7] Shi Baofeng. Research on credit rating model of small enterprises based on default pyramid principle [D]. Dalian University of technology, 2014.