

## Credit decision of small, medium and micro enterprises

Lu Yadong<sup>1</sup>, Chen Jingru<sup>2</sup>, Qin Rui<sup>3</sup>

<sup>1</sup>College of science of China Three Gorges University, Yichang, Hubei, 443002

<sup>2</sup>College of Economics and Management, Nanjing Forestry University, Nanjing, Jiangsu, 210037

<sup>3</sup>School of Mechanical and Electrical Engineering, Wuhan Donghu University, Wuhan, Hubei, 430212

**Keywords:** multi-objective planning, model checking, BP neural network, quantitative analysis, planning

**Abstract:** Based on data mining technology, this paper analyzes and processes the invoice data of input and output of various companies in recent years. Through research, grasp and use of economic laws, quantitative analysis of corporate credit risk, and establishment of a multi-target planning bank lending strategy The model realizes the optimal allocation and optimal regeneration of the credit amount, and finally passes the model test to verify the accuracy and effectiveness of the model.

### 1. Introduction

Small, medium and micro enterprises occupy an important position in my country's national economy and are an important subject of the market economy. They have played an irreplaceable role in promoting the growth of the national economy, improving the economic structure, expanding social employment, increasing fiscal revenue, and maintaining social stability. effect. There are a large number of small, medium and micro enterprises in my country, accounting for 99.7% of the total number of enterprises in the country. However, due to the small scale of production and lack of mortgage assets, credit and financial constraints have become the main obstacles to their further development, especially this year. Due to the impact of the new crown epidemic, its business development has encountered some difficulties more or less. When banks extend loans to these companies, they usually consider some factors, such as the company's bill information and the influence of upstream and downstream companies, etc. The bank uses big data by collecting and analyzing the company's settlement information, tax payment, and settlement flow. Evaluate companies to provide loans to companies with strong capabilities and stable supply-demand relations.

The use of big data to analyze the strength and reputation of a company can obtain a more complete customer puzzle, evaluate its credit risk, and provide companies with accurate strategies. This is of great significance for risk prevention and cost reduction. Commercial banks try their best to meet the capital needs of small and medium-sized enterprises in the rapid development on the basis of controllable risks.

Small, medium and micro enterprises have a great demand for funds, but banks are also very risky when issuing loans. It is difficult for them to find high-quality cooperation partners. This has caused the contradiction between the financing of small, medium and micro enterprises and bank credit. Therefore, it is very necessary to use big data to adjust the credit strategy to achieve a win-win situation between banks and small, medium and micro enterprises as much as possible.

### 2. Establishment and solution of problem one model

#### 2.1 Model establishment

Based on data analysis, the information of the 2019 statistical data on the relationship between the bank's annual lending interest rate and the customer churn rate, after data visualization and data processing, the functional relationship between the bank's annual lending interest rate and the three

types of customer churn rates of A, B, and C is as follows:

The functional relationship between the annual interest rate of bank loans and the churn rate of type A customers:

$$y_1 = -0.1716 \cdot x^{-0.683} + 1.548$$

The functional relationship between the annual bank lending interest rate and the churn rate of type B customers:

$$y_2 = -0.2966 \cdot x^{-0.546} + 1.721$$

The functional relationship between the annual interest rate of bank loans and the churn rate of category C customers:

$$y_3 = -0.5104 \cdot x^{-0.431} + 2.045$$

Among them, x represents the annual interest rate of bank loans, which are respectively represented as the churn rate of A, B, and C customers.

### 2.1.1 Establish a risk assessment model based on BP neural network

The discretization method is used to quantify the default situation, that is, use 1 to indicate that the company is in default and 0 to indicate that the company does not default. This is based on the econometric model to explain the company's default.

At this time, the company's default situation has become a discrete random variable with values of 0 and 1.

At the same time, the company's default status mainly depends on the company's operating status and the company's future development status. The company's operating status can be determined by the total amount of input invoices, the total amount of output invoices, the total number of input invoices, the total number of output invoices, and the total The tax amount reflects, and the future development of the enterprise can be reflected by the growth trend of the enterprise input value, the growth trend of the output value, the ratio of invalid invoices and negative invoices to transaction invoices, because the more frequently invalid invoices and negative invoices appear, the better they can be. The above reflects whether the company's supply and demand relationship is stable.

Take the company as an example, according to Matlab programming, some of the data obtained are shown in Table 1.

Table 1: Company data

| index                           | E1             | E2           | E3           |
|---------------------------------|----------------|--------------|--------------|
| Total amount of input invoice   | 6892693842.00  | 171736465.85 | 57520838.33  |
| Total amount of output invoice  | 6946925979.04  | 48758931.29  | 780945531.47 |
| Total number of input invoices  | 3441           | 32156        | 4561         |
| Total number of output invoices | 8110           | 12707        | 24073        |
| Proportion of voided invoices   | 3.6%           | 3.9%         | 2%           |
| Negative invoice ratio          | 2.6%           | 1.1%         | 15.8%        |
| Total tax                       | 1574484761.53  | 46518636.35  | 96282840.39  |
| Growth trend of input value     | -2074911623.90 | -18904332.79 | -13446090.65 |
| Sales growth trend              | -6873226817.34 | 9487052.75   | 686512633.63 |

With the above indicators as the input layer and the quantified default () as the output layer, a risk assessment model based on BP neural network is established. After the training of BP neural network, the results obtained by each company will be placed in the middle. There will be a few more than 0 or 1. The larger the value, the greater the bank's credit risk to the enterprise. Therefore, according to the results obtained by the risk assessment model, the risk of each enterprise is quantitatively assessed.

### 2.1.2 Quantitative corporate reputation rating

The probability distribution method is used to quantify the reputation rating. After data processing on the attachments, the proportions of the four types of enterprises A, B, C, and D in the total number of enterprises are obtained, as shown in Table 2.

Table 2: The proportion of the four types of enterprises in the total number of enterprises

| grade      | A      | B      | C      | D      |
|------------|--------|--------|--------|--------|
| Percentage | 21.95% | 30.90% | 27.64% | 19.51% |

Since the level distribution has a quasi-normal distribution, its expected value determines the position of the density function, and the standard deviation determines the extent of the distribution. Based on this, the proportion of each level is fuzzy, and then each level is assigned in the interval. The higher the value, The higher the untrustworthiness rate of the company, the interval average is used instead of the overall method to quantify the reputation level. The specific formula is as follows:

$$K_j = \sum_{k=1}^j (T_j/2) + \sum_{k=1}^{j-1} (T_{j-1}/2)$$

Among them,  $T_j$  represents the interval length of the  $j$ th level, that is, the proportion of this level. The proportion of each grade is shown in Table 3.

Table 3: The proportion of each grade and the quantitative ratio

| grade           | A    | B    | C    | D    |
|-----------------|------|------|------|------|
| Percentage      | 20%  | 30%  | 30%  | 20%  |
| Quantized value | 0.10 | 0.35 | 0.65 | 0.90 |

In order to show the relationship between the data more clearly, the data is visualized as shown in Figure 7.

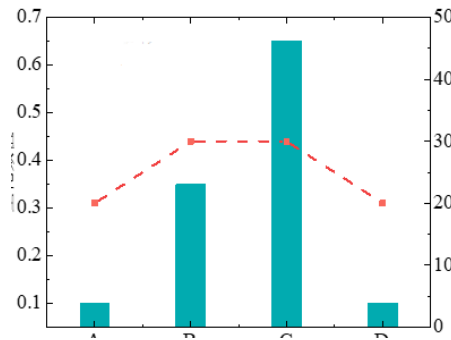


Figure 7: The proportion of each level and the quantitative ratio

From the above figure, the proportion of grades B and C is relatively high, and the proportion of grades A and D is relatively the lowest. The image satisfies the characteristics of quasi-normal distribution.

### 2.1.3 Establish a credit decision model with default risk parameters and customer churn rate

After data processing, there are two factors that affect whether a bank lends or not, namely, the bank's credit risk to the enterprise and the enterprise's untrustworthy rate. There are three factors that affect the amount of bank lending, namely the bank's total income, risk credit amount, and customer churn rate. Based on this, a "three-step" credit decision model is established.

**Step 1:** Because the bank's credit risk to the enterprise is related to the corporate reputation, the credit risk is selected to represent the bank's lending standards. According to the obtained four types of credit ratings of A, B, C, and D, the credit risk range corresponds to the credit risk range, because the bank's credit rating is Type D companies do not lend, but instead lend to Type A, B, and C companies, so companies that do not lend by banks can be preliminarily excluded.

**Step 2:** Assuming that the bank's fixed annual total credit is M, the bank needs to allocate the amount based on the credit risk value of the company to be invested (100 million 1 million). Here, we first use the planning method to describe whether the bank conducts business on companies that have reached the standard. Lending:

Use the entropy method to determine the factors that affect whether to lend: the bank assigns weights to the credit risk of the enterprise and the untrustworthy rate of the enterprise, and the weight distribution is 0.5, 0.5. Based on this, the standard of whether the bank lends is obtained, expressed by, and the calculation method is as follows:

$$l_i = \omega_1 \cdot q_i + \omega_2 \cdot p_i$$

Among them, represents the corporate untrustworthiness rate, represents the credit risk value of the first credit company, and represents their respective weights.

Used to describe the bank's credit amount to the i-th credit enterprise, the bank loan risk amount can be expressed as:

$$H = \sum_{i=1}^n z_i \times b_i \times l_i$$

Among them, represents the total number of companies waiting to lend.

Step 3: Since bank lending will have corporate defaults, in order to guarantee profits, banks not only need to invest funds in low credit risk companies, but also allocate interest rates to obtain profits. According to this, one year later, the bank's profit can be obtained as:

$$F = \sum_{i=1}^n z_i \times b_i \times r_i \times (1 - w_i)$$

Among them, represents the bank's credit interest on the i-th credit enterprise; represents the default rate of the i-th enterprise. The corresponding default rate can be obtained according to the credit rating of the i-th company. The specific results are shown in Table 4.

Table 4: Default rates corresponding to different credit ratings

| Credit rating | A | B     | C     | D |
|---------------|---|-------|-------|---|
| Default rate  | 0 | 0.026 | 0.059 | 1 |

Because banks will lose potential customers due to changes in interest rates and other factors, this is expressed in terms of customer churn rate. In the case of ensuring that the bank has a small credit risk value for the enterprise and the objective of the bank's loan income, it is also necessary to ensure that the bank has more and more potential customers.

Since data processing has determined the non-linear exponential relationship between bank lending rates and customer churn rates of various levels, the bank's mortgage rates need to be restricted at this time. The smaller the value, the lower the potential customer churn rate. Assume that the customer churn rate for unsuccessful loans is 1, so you only need to sum the customer churn rate under each lending interest rate to get the average. The function expression is as follows:

$$W = \sum_{i=1}^n (y_i / n)$$

Among them,  $y_i$  represents the customer churn rate of the i-th company when the interest rate is. If the company meets the loan demand but the bank does not lend, then the bank's churn rate for the company is 1.

Based on this, a credit decision model with default risk parameters and customer churn rate is established. Because the purpose of bank lending is to obtain income, and to reduce the amount of loan risk and reduce the customer churn rate as much as possible, the objective function can be obtained as:

$$\left\{ \begin{array}{l} F = \sum_{i=1}^n z_i \times b_i \times r_i \times (1 - w_i) \\ \min H = \sum_{i=1}^n z_i \times b_i \times l_i \\ \min W = \sum_{i=1}^n (y_i / n) \end{array} \right.$$

Among them, F, H, W respectively represent the bank's total income, the amount of bank risk credit, and the average churn rate of potential bank customers.

This is a typical multi-objective linear programming, first of which is normalized and unified dimension. When setting a single objective, the maximum solution of F is, then after normalization, it is. Similarly, if the minimum values of H and W are respectively and, they will be normalized to and.

For two secondary goals, use the square sum weighting method, also known as the virtual goal method. If equal weights are taken, the secondary goal can be written as:

$$\min S = 0.5(\overline{H}/H)^2 + 0.5(\overline{W}/W)^2$$

Then, the main goal and the secondary goal use multiplication and division, then all the goals can be combined into:

$$\max Z = \frac{F}{S} = \frac{F/\overline{F}}{0.5(\overline{H}/H)^2 + 0.5(\overline{W}/W)^2}$$

The constraints are divided into the following four parts (amount is in ten thousand units):

- 1) The credit amount must be between 10,000 yuan, namely:  $10 \leq b_i \leq 100, \quad i = 1, 2, \dots, n$
- 2) As the credit risk value is smaller, the bank trusts it more, and the interest rate is relatively low at this time. Assuming that the data at this time are arranged according to the credit risk value from small to large, then the constraints are:  $r_{i+1} \geq r_i, \quad i = 1, 2, \dots, n - 1$
- 3) Because banks have restrictions on the interest allocated by credit enterprises, namely:  $0.04 \leq r_i \leq 0.15, \quad i = 1, 2, \dots, n$
- 4) The total bank credit amount shall not exceed the fixed annual total credit amount, namely:  $\sum_{i=1}^n z_i \times b_i \leq M.$

In summary, the objective function is:

$$\max Z = \frac{F/\overline{F}}{0.5(\overline{H}/H)^2 + 0.5(\overline{W}/W)^2}$$

The constraints are:

$$s. t. \left\{ \begin{array}{l} z_i = 0 \text{ or } 1, \quad i = 1, 2, \dots, n \\ 10 \leq b_i \leq 100, \quad i = 1, 2, \dots, n \\ r_{i+1} \geq r_i, \quad i = 1, 2, \dots, n - 1 \\ 0.04 \leq r_i \leq 0.15, \quad i = 1, 2, \dots, n \\ \sum_{i=1}^n z_i \times b_i \leq M \end{array} \right.$$

## 2.2 Model solution and result analysis

### 2.2.1 Solution and result analysis of BP neural network

In the risk assessment model based on the BP neural network, the nine indicators in the previous

article are the input layer, and the quantified default () is the output layer. After the training of the BP neural network, it can be clearly seen that the model is The prediction effect of investment risk is good, the trend of the predicted value is consistent with the trend of the actual value, the value of the test sample set is close to the real value, the prediction result is very accurate, and the prediction accuracy is actually acceptable, and All of them are in between, indicating that the BP neural network training is relatively successful and has certain reference value for predicting investment risks in actual situations.

### **2.2.2 Solution and result analysis of credit decision model**

Based on the establishment of the above model and programming with Matlab, the bank's credit strategy for these companies when the total annual credit is fixed, and the annual interest rates that various companies need to pay when applying for different loan lines are solved.

The data can be seen: the larger the loan amount of the enterprise, the larger the annual interest rate. The level of interest rate has a certain relationship with the loan amount, loan time, and the credit rating of the lender. Generally speaking, the longer the loan time and the larger the loan amount, the higher the interest paid. Interest is also affected by the conditions of the company. If the company has good credit and stable funds, even if the loan is large, it can still enjoy a relatively low borrowing interest rate.

Sometimes the loan amount and the annual interest rate are not necessarily positively correlated. This is because although some companies do very large businesses, the number of invoices and the total amount of invoices are high each year, and the credit rating is relatively good. So even if some companies have higher loan lines, their annual interest rates are relatively low.

## **3. Evaluation and promotion of the model**

In this paper, a number of small models are established step by step, including the risk assessment model based on BP neural network. After neural network training, it is convenient to quantitatively estimate the risks of each enterprise. In order to unify the dimension and analyze the relationship between credit risk and corporate reputation level, after testing, it can be seen that there is a specific relationship between the two. Therefore, credit risk value is used instead of reputation rating. This test makes the model establishment easier. However, there are certain requirements for the comprehensiveness of the data, and the accuracy of the data provided in the attachment cannot be guaranteed, so some errors may be caused in the process of solving the model.

With the rapid development of Internet technology in today's society, the integration of technology and finance has become increasingly close, and big data has penetrated into the field of credit decision-making. With regard to the credit decision-making issues of small, medium and micro enterprises, due to the complexity of various types of enterprise data, which is difficult to analyze and distinguish manually, traditional credit methods alone cannot meet the needs of banks. Therefore, the emergence of intelligent credit decision-making models is crucial. The strategy should make full use of big data to reduce unnecessary manpower and financial resources. The proposed credit strategy is highly popularized, and the established model is suitable for common credit strategy problems, and the required answers can be obtained by modifying the corresponding conditions. The Internet provides a new channel for banks to communicate with customers, and a complete customer puzzle can be obtained through various data, so as to provide customers with targeted services.

## **References**

- [1] Preliminary accounting results of gross domestic product (GDP) for the first quarter of 2020, website of the Bureau of Statistics, April 2020, [http://www.gov.cn/xinwen/2020-04/18/content\\_5503803.htm](http://www.gov.cn/xinwen/2020-04/18/content_5503803.htm).
- [2] Cao Pei. Commercial bank customer credit risk assessment based on BP neural network[J]. Financial Management Research, 2018(02): 46-62.

- [3] Du Yiqian. Design and Implementation of Consumer Credit Intelligent Decision Model System[J]. China New Technology and New Products, 2020(13): 13-15.
- [4] Xie Jiaping, Chen Rongqiu. 0-1 type target programming model for optimization of product recycling treatment strategy [J]. System Engineering Theory and Practice, 2004(03): 52-57.
- [5] Xu Ge, Wang Houfeng. The development of topic models in natural language processing[J]. Chinese Journal of Computers, 2011, 34(08): 1423-1436.
- [6] Huang Hongwei. Research on time series hybrid forecasting method based on LSTM[D]. Xiangtan University, 2018.
- [7] Wang Yu. The application of MATLAB in curve fitting [J]. Science and Technology Innovation and Application, 2012(18): 8-9.
- [8] Zha Yanping, Wang Gang. Multi-objective linear programming MATLAB software solution[J]. Journal of Zaozhuang University, 2010, 27(05): 25-27.
- [9] Li Deyi and Li Ming, Mathematical Modeling (Second Edition), Science Press, February 2016, pages 39-49.
- [10] Chen Dongyan, Liu Fengqiu and Niu Ben, Mathematical Modeling (Second Edition), Science Press, January 2014, pages 188-208.
- [11] Li Xin, Mathematical Modeling in MATLAB, Tsinghua University Press, December 2017, pages 355-392.
- [12] Fu Wenli and Liu Gang, MATLAB Programming Guide, Tsinghua University Press, December 2017, pages 61-95.