

Credit Risk Assessment of Small, Medium and Micro Enterprises Based on The Improved Logical Regression Model of Gradient Descent Method

Congcong Wu*

School of Economics and Management, Northeast Electric Power University, Jilin Province, China

* Corresponding Author Email: 2636008389@qq.com

Abstract. Due to the weak strength of small, medium and micro enterprises, banks need to judge the credit risk of the enterprise based on the company's past credit history and the company's invoice information, etc., and then establish a reasonable and effective quantitative credit risk model. First, through the analysis and preprocessing of the data, this paper extracts six indicators such as marketing profit margin and invoice invalidation rate, and comprehensively evaluates the credit risk of the enterprise from the three aspects of enterprise scale, supply and demand stability, and credibility, and establishes a corporate credit risk evaluation system. Then, a logical regression model is constructed to predict the probability of compliance of the enterprise, a loss function is formulated to describe the deviation between the prediction result and the classification result, and a stochastic gradient descent algorithm is used to obtain the optimal parameter value. Finally, the optimized logical regression model is applied to the credit risk assessment of 123 companies. The results show that the stronger the company's strength, the more stable the supply and demand relationship, and the higher the reputation level, the less likely its credit default risk is.

Keywords: Small, medium and micro enterprises; credit risk; probability of enterprise compliance; improved logical regression model based on gradient descent method.

1. Introduction

1.1. Background of the problem

Small, medium and micro enterprises have a positive effect on enhancing the economic vitality of our country. However, the small size of small, medium and micro enterprises, the lack of collateral, and the large risk fluctuations make it difficult for them to obtain bank loans, which seriously restricts the healthy development of small, medium and micro enterprises. Therefore, banks must first evaluate the credit risk of small and medium-sized enterprises in terms of their size and credit.

1.2. Literature Review

In terms of credit risk assessment for small and medium-sized enterprises, many scholars have proposed many methods. Chunxiu Zhao [1] et al. combined the rough set and BP neural network to evaluate small and medium-sized enterprises, and focused on the introduction of mesoscopic factors affecting customer credit in customer credit risk assessment, and also used AHP to study the risk management of small and medium-sized enterprises and put forward corresponding suggestions. Zepeng Song [2] et al. used qualitative analysis method and independent sample T-test method to test for significant differences, and then used the neural network toolbox to construct a commercial bank credit risk identification model. And through the introduction of samples, the model is trained and the law of identifying corporate credit risks is discovered. Meng Zhang [3] and others used the principal component analysis method and the expert scoring method to screen and score financial and non-financial indicators, and used the Logical model to construct a credit risk evaluation model.

Inspired by Shangmiao Wen [4] using the Logit model to establish a credit evaluation model for small and medium-sized enterprises, this paper proposes to use a gradient descent logical regression model to quantify the credit risk of enterprises.

2. Data preprocessing

The data in this article is derived from Question C of the 2020 National Mathematical Modeling Competition for College Students [5]. After observation and analysis, we need to preprocess the data. Mainly includes the following points:

- (1) Filter and delete enterprise data with a reputation level of D.
- (2) Filter and delete invalid invoices.
- (3) Filter and delete the invoice data of most enterprises with a small number of invoices in a certain year.

3. Establishment of evaluation index system

Through analyzing the database and reading relevant literature, this article transforms the quantitative analysis of the credit risk of 123 enterprises into the establishment of a default risk evaluation model for 123 enterprises, and extracts a total of 6 characteristics: reputation rating, total input price tax, total sales price tax, profit, sales profit margin and invoice invalidation rate, and can measure the credit risk of small and medium-sized enterprises from the three aspects of enterprise size, supply and demand stability and reputation level. The specific evaluation indicators are as follows:

3.1. Reputation rating (RR)

Credit evaluation is the analysis and research of the credit risk factors of the evaluated target, so as to obtain a comprehensive assessment of the solvency and willingness of the evaluated target to repay debts.

3.2. Total input price tax (TIPT)

The total input price tax refers to the sum of the payment amount and tax payable by the enterprise when purchasing goods or receiving services. The formula is as follows:

$$TIPT = PA + TPOP \quad (1)$$

Among them, PA is the amount actually paid by the enterprise when purchasing goods or receiving services; TPOP is the amount of value-added tax that the enterprise needs to pay when purchasing goods or receiving services according to national regulations.

3.3. Total sales price and tax (TSPAT)

The total sales price and tax refers to the sum of the taxes and sales involved in the process of selling products or providing services. It is composed of two parts: S (Sales) and VAT (Value added tax). The formula is as follows:

$$TSPAT = S + VAT \quad (2)$$

3.4. Profit (P)

Profit refers to the amount remaining after subtracting costs and other expenses from the income obtained by an enterprise from the sale of goods or the provision of services. The formula is as follows:

$$P = TSPAT - TIPT \quad (3)$$

3.5. Sales profit margin (SPM)

Sales profit margin refers to the proportion of profits realized by an enterprise from its sales revenue after selling goods or services. Its formula is as follows:

$$SPM = \frac{P}{TSPAT} \quad (4)$$

The numerator is profit, and the denominator is sales revenue.

3.6. Invoice invalidation rate (IIR)

The invoice invalidation rate refers to the proportion of the number of invoices that need to be invalidated for various reasons in the invoices issued by the enterprise to the number of all invoices that have been issued. The specific formula is as follows:

$$IIR = \frac{NOII}{TNOI} \quad (5)$$

Among them, NOII is the number of invalid invoices of the enterprise, and TNOI is the total number of invoices of the enterprise.

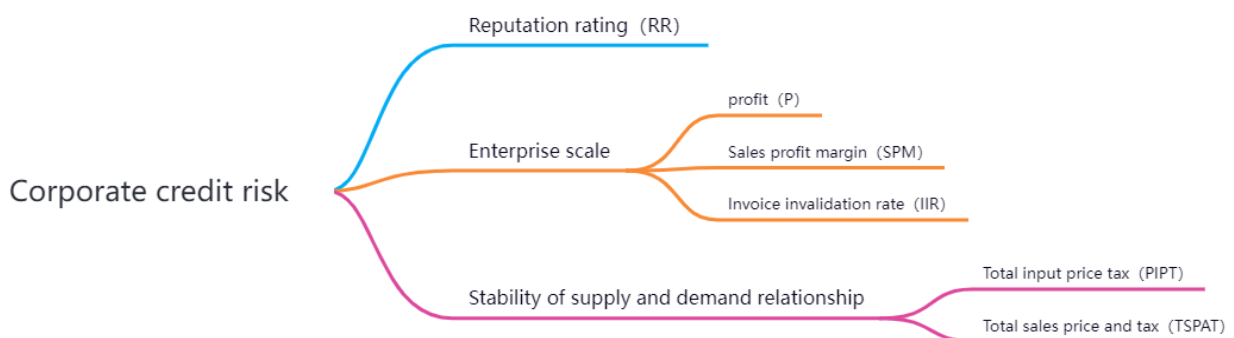


Figure .1 Corporate credit risk evaluation index system Diagram

4. Credit risk assessment of small, medium and micro enterprises

4.1. Gradient descent logical regression model establishment

4.1.1 Establishment of logical regression model

Logical Regression is suitable for solving the regression problem where the dependent variable is a categorical variable. The definition of the logical regression model is as follows [6]:

$$\begin{cases} y = \frac{1}{1+e^{-z}} \\ z = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_6 x_6 \end{cases} \quad (6)$$

Where y refers to the probability of the company's compliance to be observed, and finally the probability that the value of the company's compliance rate is between $[0,1]$ is obtained. x_1, x_2, \dots, x_6 refer to the six characteristics that affect whether the enterprise complies with the contract, and θ is the weight parameter.

4.1.2 Loss function establishment

In order to minimize the error between the prediction result and the real result, the minimized mean square error is used as the loss function of the logical regression, so that the optimization goal of the prediction problem in this study is [7]:

$$\min_{\theta} y = \frac{1}{2m} \sum_{i=1}^m (y^i - y_{\theta}(x^i))^2 \quad (7)$$

Where m is the sample size. $y_{\theta}(x^i)$ is the prediction result obtained by training the i -th sample; y^i is the true result (label) of the i -th sample. To construct an accurate prediction model of whether an enterprise is complying with the contract, it is required to solve the optimal value of the parameter θ in formula (13).

4.2. Model solving

The Mini—batch Gradient descent method (MBGD) is used to solve the model. Its mathematical expression is:

$$\begin{cases} \frac{\partial y}{\partial \theta_j} = -\frac{1}{M} \sum_{i=1}^M (y^i - y_{\theta}(x^{(i)})) x_j^{(i)} \\ \theta_j := \theta_j + \eta \frac{1}{M} \sum_{i=1}^M (y^i - y_{\theta}(x^{(i)})) x_j^{(i)} \end{cases} \quad (8)$$

Where $x_j^{(i)}$ is the j-th input component of the selected sample in the i-th generation, and η is the learning rate. The parameter θ is updated by repeated generation selection according to formula (6) until the error between the two iterations is lower than a certain fixed threshold value. The update ends. The θ value obtained at this time is the optimal parameter. Substitute it into formula (6) to obtain a prediction model of whether the enterprise is complying with the contract. Only need to enter the 6 characteristic data of any enterprise ($x_1, x_2 \dots x_6$) into formula (6) to accurately predict the probability of its enterprise's compliance with the contract, and provide the bank with a basis for credit decision-making.

4.3. Experimental results and analysis

Herein, the use of gradient descent logistic regression model to predict each of the small and micro enterprises corporate compliance rate, and the obtained confusion matrix heat map and model the results of the assessment table.

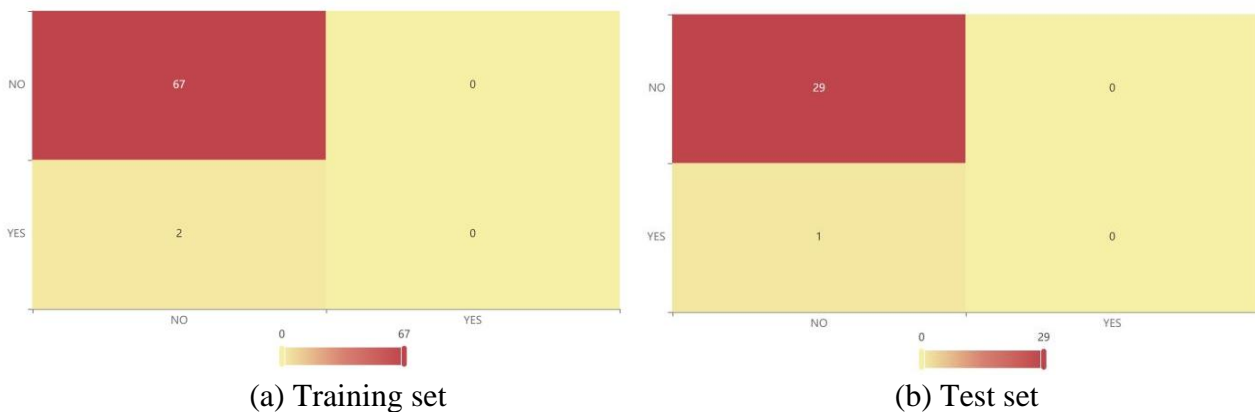


Figure. 2 The confusion matrix heat map

From Figure 2(a), we know that in the training set, there are 69 data of whether the enterprise defaults, of which 67 "no" data are correctly predicted, and only 2 "yes" data are incorrectly predicted. Therefore, the system is better at distinguishing between "yes" and "no".

From Figure 2(b), we know that in the test set, there are 30 data of whether a company is in default or not, and 29 "no" data are predicted correctly, while 1 "yes" data is predicted incorrectly. Therefore, the system is more effective in distinguishing between "yes" and "no" tests.

Table .1 Model evaluation results

	Accuracy	Recall rate	The exact rate	F1
Training set	0.971	0.971	0.943	0.957
Test set	0.967	0.967	0.934	0.95

According to Table .1, the classifier performs well on the training set and the test set. The accuracy rate, recall rate, accuracy rate, and F1 value are all high, indicating that the classifier can effectively classify and predict, and maintain consistent performance on different data sets.

Table .2 Enterprise compliance probability prediction results

Enterprise	Probability of compliance	Enterprise	Probability of compliance
E1	1.00000	E84	0.99442
E2	1.00000	E85	0.99442
E4	0.99442	E88	1.00000
E5	1.00000	E90	1.00000
E6	1.00000	E91	1.00000
E8	1.00000	E94	0.99442
E10	1.00000	E95	1.00000
E11	1.00000	E96	1.00000
E13	1.00000	E98	0.99442
E14	0.99442	E104	0.99442
E15	1.00000	E106	0.99442
E16	1.00000	E110	1.00000

Due to limited space, this article only shows the data of the first 12 companies and the last 12 companies, including the corporate code and the probability of corporate compliance predicted by the model. The closer the probability is to 1, the greater the probability that the company will abide by the contract and repay the loan, and the smaller the risk of default. Banks can lend to them. For example, the predicted compliance rate of Enterprise E4 is 0.99442, which is very close to the 100% compliance rate and the level of zero credit risk. At this time, banks can lend and enlarge the amount.

5. Summary

This paper conducts a quantitative analysis of the credit risks of 123 enterprises, and transforms the problem of quantitative corporate credit risk into the problem of corporate compliance probability. From the above analysis results, it can be seen that companies with high probability of compliance have the following characteristics: the operating profit of the enterprise is huge, the sales profit margin is also higher, and the smaller the invoice invalidation rate, the stronger the profitability. Secondly, their supply and demand relationship is often relatively stable. On the one hand, their total input price tax is relatively large, which can effectively drive down the price of raw materials; on the other hand, their total sales price tax is large, they have good sales channels, and the supply and demand relationship between enterprises is very stable. Finally, their reputation rating also remains high. This is related to their corporate culture and the formulation of organizational performance.

References

- [1] Zhao Chunxiu. Research on Credit Risk Analysis and Management of Commercial Banks [D]. Tianjin University, 2008.
- [2] Song Zepeng. Research on the credit risk identification model of small and medium-sized enterprises in commercial banks [D]. Capital University of Economics and Business, 2016.
- [3] Zhang Meng. Research on the credit risk evaluation of Small and medium-sized enterprises in China's commercial Banks [D]. North China University of Technology, 2019.
- [4] Shang Miao. Credit risk assessment and management of small and medium-sized enterprises based on logit model [J]. Times Finance, 2018, No.694(12):210+220.
- [5] Questions for the 2020 National Mathematical Modeling Competition for College Students. http://www.mcm.edu.cn/html_cn/node/10405905647c52abfd6377c0311632b5.html, 2021-01-22
- [6] Lan Jun, Yan Guangle. Research on credit risk of small and medium-sized enterprises from the perspective of social capital [J]. China's circulation economy, 2019, 33(05):111-119.
- [7] Li Peng, Min Hui, Qu Haoyu, etc. Research on the prediction of the incidence of ischemic stroke based on the logical regression model [J]. Journal of Medical Informatics, 2020, 41(06):28-32.