Research Article

Decision Analysis of Multifactor Credit Risk Based on Logistic Regression and BP Neural Network

Qiang Tang and Wen Yu Shi

School of Big Data and Artificial Intelligence, Anhui Xinhua University, Anhui Hefei 230088, China

Correspondence should be addressed to Wen Yu Shi; shiwenyu@axhu.edu.cn

Received 3 March 2022; Accepted 9 June 2022; Published 1 July 2022

Academic Editor: Muhammad Faisal Nadeem

Copyright © 2022 Qiang Tang and Wen Yu Shi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Small, medium, and micro enterprises play an important role in the development of the national economy and are of great significance in promoting technological innovation, relieving employment pressure, facilitating people’s lives, and maintaining social stability. But in China, small, medium, and micro enterprises generally exist in the phenomenon of “financing difficulties.” Therefore, we need to find a method to forecast its credit risk. By using Python, SPSS, and other software, based on a two-component logistic regression model, assisted by multievaluation model and supported by game theory, this paper establishes an innovative comprehensive credit risk assessment model for small, medium, and micro enterprises.

1. Introduction

With the development of China’s economy and the diversification of the structure, the economic power of domestic small and medium-sized enterprises has an indispensable and important force, and plays an important role in employment absorption, social stability, and economic vitality, and is the main contributor to domestic GDP. Therefore, promoting the development of small and medium-sized enterprises has economic and political significance. Promoting the development of small- and medium-sized enterprises not only depends on the macro control of national policies but also needs the support of actual capital, especially financial support. However, in real life, due to the relatively small scale of small, medium, and micro enterprises and lack of mortgage assets, banks usually give preferential interest rates to enterprises with high reputation and low credit risk according to credit policies. The bank first evaluates the strength, reputation, and credit risk of the small, medium, and micro enterprises and then decides whether to make loans and credit strategies such as the loan limit, interest rate, and term according to the credit risk factors.

Li Huaming (2004) and Xue Qinghai (2004) et al. believe that the main reasons for the financing difficulties of private enterprises are small scale, information asymmetry between banks and enterprises, and excessive information cost and supervision cost of financial institutions’ loans to SMEs. Lin Yifu (2001) believes that “the fundamental reason for financing difficulties of private SMEs is that SMEs have low operation transparency and imperfect financial system. Shanghai Research Group of the People’s Bank of China (2001) believes that the financing obstacles of SMEs are their low credibility, the existing credit service system of financial institutions is difficult to adapt to the financing characteristics of SMEs, and the social financing support system for SMEs has not been established. Yang Et al. (2000) ascribe the restrictive factors of SMEs’ financing to their financial irregularities, difficulty in mortgage and guarantee, industrial attributes, and ownership concepts. Chen Dongsheng (2000) believes that there are conceptual barriers, credit barriers, guarantee barriers, information barriers, and cost barriers in SME financing.

By combing and comparing the successful assessment of enterprise credit risk by domestic and foreign researchers, we can conclude that domestic research on credit decision-
2 Mathematical Problems in Engineering

2.1 Analysis of Question 1.

2.2 Analysis of Question 2.

2.3 Analysis of Question 3.

3. Problem Solving

3.1 Solve Question 1

3.1.1 The AHP Model. In this model, APH was used to establish the time-like structure describing the internal features or feature independence of the system. By comparing the relative importance of the two indicators, the relative importance of the lower related elements of the judgment matrix A of the upper elements were constructed,
and the importance order of the layer contact elements and the values of the layer contact elements were calculated. Then, the maximum characteristic root $\lambda_{\text{max}}$ of the judgement matrix is obtained through calculation, the normalized feature vector $W$ corresponding to the maximum characteristic root is obtained, and its component $W_i$ is the corresponding weight.

### 3.1.2. Comprehensive Model of Fuzzy Evaluation

In the evaluation work, the importance of each factor is different, so each factor $U_i$ is given a weight $A_i$, and the fuzzy set of weight set of each factor is represented by $A:A_i (A_1, A_2, ..., A_m)$. In order to evaluate the strength of enterprises, we need to choose reasonable indicators and parameters and, according to the above method, construct the corresponding matrix.

### 3.1.3. Establishment of the Logistic Model

Logistic regression is used to deal with regression problems in which the dependent variable is a categorical variable, such as dichotomy or binomial distribution problems and multi-categorization problems [3]. In fact, it belongs to a classification method. The relationship between the probability of dichotomous problems and independent variable is

$$P(y = 1 | x) = \frac{1}{1 + e^{-\theta^T x}}.$$  \hspace{1cm} (1)

The domain of the function is all the real numbers, its range is (0, 1), and its result has natural probabilistic significance.

For variables of type 0–1, the probability distribution formula of $y=1$ is defined as follows:

$$P(y = 1) = \rho.$$  \hspace{1cm} (2)

The probability distribution formula of $y=0$ is defined as follows:

$$P(y = 0) = 1 - \rho.$$  \hspace{1cm} (3)

The expected value formula of the discrete random variable is as follows:

$$E(y) = 1 * \rho + 0 * (1 - \rho).$$  \hspace{1cm} (4)

The linear model is used for analysis, and the formula transformation is as follows:

$$p(y = 1 | x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n.$$  \hspace{1cm} (5)

However, since probability $P$ and dependent variable are often nonlinear, we introduce logit transform to make $\text{logit}(p)$ linearly correlated with independent variable and establish the logistic regression model as follows:

$$\text{logit} (p) = \ln \left( \frac{p}{1 - p} \right) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n.$$  \hspace{1cm} (6)

That is,

$$P = \frac{1}{1 + e^{-(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n)}}.$$  \hspace{1cm} (7)

After obtaining the required Sigmoid function, it is only necessary to fit the $n$ parameters $\theta$ in the equation as in the previous linear regression.

The above two models construct the set of variables that affect the credit risk of enterprises, and the probability of the occurrence of credit default of enterprises determines the value of variables in the set of variables. Therefore, in order to predict the probability of enterprise default, we used a logistic regression model to conduct research. Since there are only two values, default and timely repayment, we established a binary logistic regression model to predict the default probability.

A logistic regression model is a nonlinear prediction model for the occurrence probability of multivariable events [4]. It takes multiple indicators of the event as its independent variables, and there are only two values of the event. The occurrence of the event is marked as 1, the non-occurrence of the event is marked as 0, and the occurrence probability of the event is marked as $P$.

After building the initial model, we need to discuss the regression value of $P$. The closer it is to 1, the greater the probability of enterprise credit violation is. In addition, we will take 0.5 as the limit. If the value is greater than 0.5, it will be 1; that is, the default is judged; otherwise, there is no default.

Through the establishment of multiple logistic regression model, according to the analytic hierarchy process (AHP) and fuzzy evaluation method, the index with higher weight is regarded as the index with higher contribution rate. In our logistic regression analysis, the SPSS software is used and the method of forward-backward mixed stepwise regression is adopted. The inconsistent deleted variables are entered into the back branch step by step, and the iterative calculation is performed. By fitting the curve, we find that the Pearson index is greater than the significant level. The specific data are shown in the following Table 1.

### 3.2. Solve Question 2

This problem will adopt the BP algorithm (error backpropagation algorithm) and the LM algorithm.

#### 3.2.1. Establish BP Neural Network Model

Artificial neural network model is a kind of bionic network dynamics model, which has self-organization, self-adaptation, and self-learning functions [5]. According to the working principle of biological nervous system, artificial neurons (basic units) of many nodes are connected to form a network in certain rules or ways to simulate and show the overall behavior of the system [6]. Backpropagation (BP) neural network model (as shown in Figure 1) is the most widely used model in artificial neural network.

After learning all the samples and averaging the errors of each sample, this algorithm can be used as the correction
error, which can significantly reduce the number of corrections and speed up the convergence. Specific algorithm is as follows:

1. Initialization, all correction parameters clear.
2. The correction error of learning mode is calculated.
3. Error accumulation, update the learning mode.
4. Judge whether all learning modes are completed. If the skip is complete, go to Step 7.
5. Calibration of connection weights and cumulative learning times.
6. Continue whether the error meets the requirements. If the error does not meet the requirements, return to Step 2.
7. End the program.

The attached invoice data training network was used to input the attached data into the established BP neural network as test data, and the predicted value of the credit rating was output and compared with the actual value to evaluate the prediction effect of the model.

The mathematical model constructed based on the artificial neural network represented by BP neural network can realize highly parallel distributed processing by relying on the network connection weight coefficient and has the ability of associative memory, self-organization, and self-learning [7]. Through training and learning, it can approximate any nonlinear mapping and has high prediction accuracy for the development trend of reputation. BP neural network fitting and prediction results are shown in Table 2.

3.2.2. LM Algorithm. LM algorithm is an iterative algorithm that can be used to solve least square problems. It can be thought of as a combination of the fastest descent method and Gauss-Newton method. When the current solution is far away from the optimal solution, the algorithm is closer to the fastest descent method: the direction of the fastest gradient descent accelerates the iteration speed; when the current solution is close to the optimal solution, the algorithm is close to GN and converges rapidly [8]. In this problem, Levenberg-Marquardt algorithm is selected to train the model [9–11].

According to the above theory, both machine learning and deep learning require a large number of samples. If the
sample size is too small, on the one hand, it takes too long, and on the other hand, it may lead to over-fitting [12]. Therefore, the method adopted here is the cross-validation method, and the data are evenly divided into 10 data blocks, one of which is taken as the test set and the other 9 as the training set, so that 10 similar data can be obtained for training.

For the data adopted, it is a one-dimensional simple data. In addition to the last label data, five indicators are obtained through screening as a vector \( (A, B, C, D, \text{and} e) \), and a label \( F \) is added. The credit rating labels to be obtained are \( A, B, C, D \). So let’s say 4,3,2,1.

Set the layers and model functions of the neural network matrix. Here, kerAS, a well-known library of deep learning, was used to build the model [13]. After other data processing, the samples were divided into training sets and test sets. Add input layer and output layer to build deep learning network. Sigmoid function is used to input data in the input layer, and Softmax function is used in the output layer. The Sigmoid function is used to standardize the output into a pattern output that can be substituted into our neural network. The Sigmoid function can be used to transform the output into a pattern output with probability. We take the label with the highest probability corresponding to the label we need. Since the matrix composed of five variables is applied, the matrix of our first layer must have at least five rows, and the label of the output is a four-row probability matrix, which requires the matrix of the previous layer of the output layer to be four rows. Represented by vectors \((1, 5), (5, 3), (x, 4), (4, 1)\), this is the most basic number of layers. At this time, there are fewer layers, and the training speed of the trainer is faster [14].

In the first training, random number method is used to obtain the parameters in the hypothesis matrix and then continuous self-learning through neural network.

Through the above machine learning and deep learning, we use enterprise-related data as a training set and train the neural network to simulate the data by taking default and other data as input and reputation as output, so as to obtain the reputation rating of the enterprise. The bank’s expected profit depends on the interest rate the bank sets for its pledged loans and the value of SME pledged inventories (determined jointly by product quantity and wholesale price). If the interest rate of the bank’s inventory pledge loan is too high, it will help to improve its profit margin, but the small and medium-sized enterprises will reduce the number of products ordered, thus reducing the bank’s expected profit. If the interest rate of bank’s inventory pledge loan is too low, SMEs may overorder products, and banks will face great credit risks without guarantee from core enterprises. Therefore, banks require core companies to provide repurchase guarantees, through which banks let core companies share some of the credit risks they face. At this time, the core enterprises have the motivation to influence the ordering quantity decision of SMEs and the value of pledged inventory through the wholesale price tool, so as to reduce the risks faced by themselves and banks. In such a credit mechanism design, the loan interest rate ensures the profit margin of the bank’s credit funds, while the repurchase guarantee requirement shares the risks that the bank may face. Therefore, the bank helps solve the financing problems of small and medium-sized enterprises by improving the credit evaluation level of the inventory pledge financing of small and medium-sized enterprises.

3.3. Solve Question 3. The impact of nonoperational risks on enterprises is mainly the reduction of expected income, but still need to maintain production, so credit risks will definitely rise. For banks, this is a time to rein in credit risk more tightly, while channeling money to industries or businesses that need it more, such as information technology and finance, which have maintained growth this year (source: STATISTICS Bureau) [15].

According to our hypothesis, the epidemic will not reduce the creditworthiness of enterprises, but only increase their credit risk. Therefore, in the case of special situations, we need to increase indicators of the impact of the epidemic on the industry based on the strength and reputation of enterprises, evaluate the credit risk of enterprises, and finally formulate policies.

We still use the fuzzy comprehensive evaluation method. The factor set of enterprise credit risk assessment is \( U = (U1, U2, U3) \), which, respectively, represent the enterprise strength, enterprise reputation, and the impact of the epidemic on the industry. The evaluation set is \( V = (V1, V2, V3, V4) \), which, respectively, represent \( A, B, C, \) and \( D \). The single-factor fuzzy evaluation is carried out for each enterprise sample, and the evaluation matrix is obtained. Finally, determine the weight distribution as \( A = (0.4, 0.4, 0.2) \). Finally, the fuzzy factor vector A on U is changed into the fuzzy vector B on V through the fuzzy change, and the final grading result of enterprise credit risk assessment is obtained.

We substitute the rating results into the ultimate comprehensive mSMES risk assessment model and give the bank’s credit adjustment strategy when the total credit is fixed.

4. Conclusion

At present, there are a lot of research results on the credit risks of small, medium, and micro enterprises, many of which are inferred by using model prediction. BP neural network and logistic regression have great advantages in intelligent learning, reasoning, and prediction, so they are technically feasible.

But when it comes to actually solving the problem, you might also want to consider the following:

1. The risk assessment level and the probability of repayment on time are not absolutely accurate, and there may be a 5% error, so the impact on the final result can be considered.

2. It may not be scientific for banks to only look at the prospects within two years. Under different interest rates, the years can be extended, and the local optimal may not be the overall optimal. There are more ways to earn more money under more conditions.
In essence, all the strength, reputation, and risk assessment of small and micro enterprises are closely related. Although this paper makes a separate analysis, when risks rise, other factors will indeed be affected, requiring more complex considerations.

In view of the shortcomings mentioned above, further optimization of the quantitative method of discrete data in the index system and the processing of index data can be considered in the subsequent research. Of course, the financing difficulty of small, medium, and micro enterprises is a worldwide problem. If we want to solve this problem well, the country and banks need to provide policy support to support small, medium, and micro enterprises and solve their financing difficulties.

Data Availability
The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest
The authors declare that they have no conflicts of interest.

Authors’ Contributions
This work was equally contributed by all authors.

Acknowledgments
This paper was supported by Quality Engineering Project of Anhui Province 2016ckj101, 2019mooc351, 2020jyxm0782, 2019kyqd01, kyt202203.

References