Financial Engineering and Risk Management (2021) 4: 78-81
Clausius Scientific Press, Canada

DOI: 10.23977/ferm.2021.040509
ISSN 2523-2576

Research on Bank Credit Model Based on Decision Tree

Tiantian Wang

School of Economics and Finance, International Studies University, Shaanxi, Xi 'an, 710100

Keywords: decision tree, credit risk, decision optimization model

Abstract: With the rapid development of small and medium-sized enterprises, banks need to establish a sound credit risk identification system. This paper makes a quantitative analysis on the credit risk of 302 enterprises without credit records, and gives the bank's credit strategy for these enterprises when the total annual credit amount is 100 million yuan. Using ID3 decision tree learning algorithm to predict the credit risk level of 302 enterprises, the credit strategies of 302 enterprises are obtained.

1. Introduction

Due to the high repayment risk of small, medium and micro-sized enterprises, banks are often reluctant to lend. In this context, in order to ensure the survival and further development of SMEs, banks need to establish a more scientific credit risk assessment system, [1] so as to provide growth space for high-quality SMEs. Traditional credit risk assessment methods include credit rating method and credit scoring method. [2] However, these methods have common problems, i.e. the quantitative standards are not uniform, etc.

2. Credit Rating Prediction Model based on Decision Tree method

By calculating the gross profit margin, return on net assets, total assets growth rate and net assets growth rate of each enterprise in 2017, 2018 and 2019 respectively, the future fluctuation trend of each index and its influence on enterprise risk assessment can be roughly inferred through the change of index data of each enterprise in different years. As can be seen from Figure 1, the net profit growth rate of different enterprises is significantly different, followed by the total assets growth rate, while the difference between gross profit margin and return on net assets of different enterprises is not significant. Therefore, the gross profit margin of sales is determined as 0.25, the return on net assets is 0.5, the return on total assets is 0.75, and the return on net profit is 1.

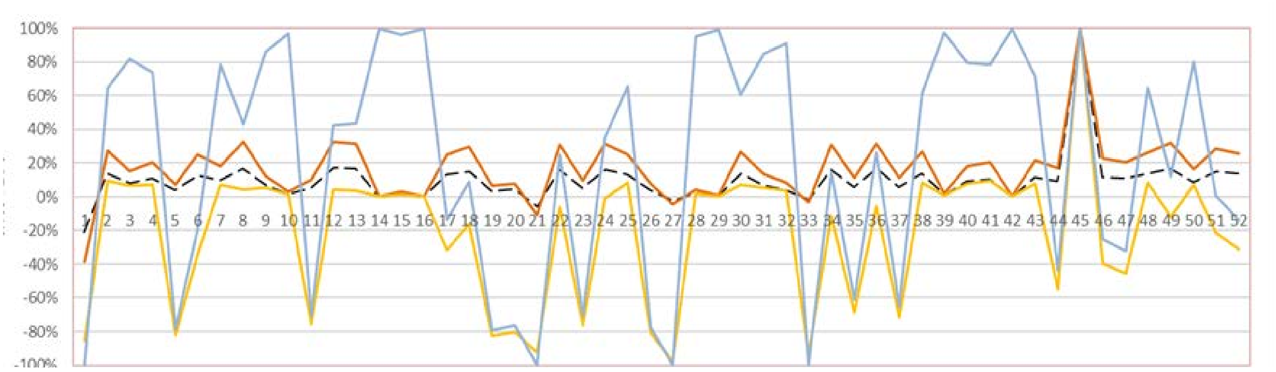

Figure 1: Changes in Financial Indicators of Different Enterprises

The elements between adjacent levels are related, and a simulation consistent judgment matrix is constructed, which is expressed as follows:

$$R=\begin{bmatrix} r_{11} & r_{21} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix}$$

And the requirements are met:

$r_{ii}=0.5, i=1,2,\cdots,n$

$r_{ij}=1-r_{ji}, i,j=1,2,\cdots,n$

$r_{ij}=r_{ik}-r_{jk}, i,j,k=1,2,\cdots,n$

The financial indicator simulation consistent judgment matrix thus constructed is

$$\begin{bmatrix} 0.5 & 0.75 & 0.4 & 0.90 \\ 0.25 & 0.5 & 0.2 & 0.65 \\ 0.55 & 0.8 & 0.5 & 0.95 \\ 0.1 & 0.35 & 0.05 & 0.5 \end{bmatrix}$$

Similarly, the simulated consistent judgment matrix of available non-financial indicators is

$$\begin{bmatrix} 0.5 & 0.8 & 0.4 & 0.75 & 0.65 \\ 0.2 & 0.5 & 0.1 & 0.45 & 0.35 \\ 0.6 & 0.9 & 0.5 & 0.85 & 0.75 \\ 0.25 & 0.25 & 0.15 & 0.5 & 0.4 \\ 0.35 & 0.65 & 0.25 & 0.6 & 0.5 \end{bmatrix}$$

The decision tree is based on the probability of occurrence of various situations. By constituting the decision tree to obtain the probability of the net present value greater than or equal to zero, the risks of the project can be used to evaluate the risk of the project.

·Construction of the concept of information entropy

Information entropy is a measure of uncertainty for random variables. Assuming that the proportion of the k-th sample in the current sample set d is p, the information entropy is defined as (k= 1,2,3,4, ... |y|), and the information entropy of d is defined as p^{k}

$$Ent(D) = -\sum_{k=1}^{|y|} p^{k} \log_2 p^{k}$$

Among them, the lower the value of $Ent(D)$, the higher the purity of D.

·Construction of the concept of information gain

The information gain is the difference between the information entropy of the parent node and the total information entropy of all nodes below. However, the simple summation of the total information entropy of the sub-nodes needs to be modified before the summation.

Assuming that the discrete property Aahas a V pi possible value $\{a^1, a^2, ..., a^V\}$, if the sample set D is used to divide the sample set D, then the V-bit direct point is generated, where the word V is directly The sample containing all the value a^V on the attribute a is recorded as D^V. We can calculate the information entropy of D^V, and considering the number of samples contained in different branch points, give the branch point $\frac{|D^V|}{|D|}$, the more the amount of the number of branch nodes is the most, so that the information gain obtained by dividing the sample set D can be calculated:

$$Gain(D, a) = Ent(D) - \sum_{v=1}^{V} \frac{|D^V|}{|D|} * Ent(D^V)$$

Among them, the greater the information gain, the greater the purity improvement obtained by dividing by attribute a. The ID3 decision tree learning algorithm uses information gain as the criterion to divide attributes and generate the final result of decision tree calculation.

·The operation result based on ID3 decision tree learning algorithm

The following table shows the credit risk rating of the enterprise derived from the relative data of the forecast processing:

Table 1: Risk Rating of Some Enterprises

Enterprise code	Enterprise name	Credit risk rating
E127	Self-employed E127	A
E295	* * * Electrical Equipment Manufacturing Co., Ltd.	B
E400	* * * Office Supplies Operation Department	C

3. Bank Credit Strategy

Based on the above risk forecast, the bank credit strategy model is established immediately when the bank's total annual credit is RMB 100million [3]:

$$B_1 + B_2 + B_3 = 1(100\text{million yuan})$$

$$max \quad Z = (1 - P_A - l_A) \cdot B_1 \cdot r_1 + (1 - P_B - l_B) \cdot B_2 \cdot r_2 + (1 - P_C - l_C) \cdot B_3 \cdot r_3 + \cdots + (1 - P_I - l_i) \cdot B_j \cdot r_j$$

$$s.t.\begin{cases} 0 \le P_I \le 1 \\ r_1 = 0.121 l_A + 0.020 \\ r_2 = 0.126 l_B + 0.022 \\ r_3 = 0.125 l_C + 0.023 \\ 0 \le l_i \le 1 (i = A, B, C) \\ 4\% \le r_j \le 15\% (j = 1, 2, 3) \\ 10 \le B_j \le 100 (j = 1, 2, 3) \\ B_1 + B_2 + B_3 = 1 \end{cases}$$

Based on the credit risk assessment results of 302 enterprises, and combined with the sales of each enterprise, we selected the data of 3 of them and applied the credit strategy model to estimate their loan amount, annual interest rate and interest rate preference range. The results are shown in the following table:

Table 2: Estimates of Credit Strategies of 3 Enterprises

Enterprise code	Enterprise name	Credit risk level	r_j	B_j	Interest rate margin
E127	Self-employed E127	A	7.4%	30.87 million yuan	1.5%
E295	* * * Electrical Equipment Manufacturing Co., Ltd.	B	9.3%	15.65 million yuan	0.25%
E400	* * * Office Supplies Operation Department	C	13.6%	0.85 million yuan	0

4. Conclusion

In this paper, the decision tree is used to predict the reputation ratings of 302 enterprises with the help of the historical data of 123 enterprises and the corresponding reputation ratings and the historical data of 302 enterprises. Then the credit risk assessment model is used to analyze the credit risk of 302 enterprises, and the multi-objective optimization model is used to obtain the best credit strategy.

References

[1] Xie Liang; Li Ning; Credit risk of private enterprises and preventive measures [J]; Journal of Inner Mongolia University of Finance and Economics; Issue 01, 2013

[2] Xing Zhe; Song Zhiqing; ; Analysis of credit advantages of small and medium-sized banks under credit market segmentation [J]; Financial forum; Issue 07, 2010

[3] An Weigang. Multi-objective optimization method