Journal of Computers Vol. 32 No. 6, 2021, pp. 159-167
doi:10.53106/199115992021123206013

Credit Index Weight Model and Application of Multi-classification Support Vector Machine Based on Particle Swarm Optimization

Zhan-Jiang Li*, Qin-Jin Zhang, Tong-Tong Wang

College of Economics and Management, Inner Mongolia Agricultural University, Hohhot, China
{lizhanjiang582, zqj1575212, wangtongtong4009}@163.com

Received 1 July 2021; Revised 1 August 2021; Accepted 2 August 2021

Abstract. In the current credit evaluation environment with small default samples and unbalanced default status, it is a meaningful research work to explore the weight system of credit indicators with high discriminative power and high precision, which is relatively lacking in the existing researches. First by loss function to enterprise's credit status is divided into high and low default default and no default three categories, and then support vector machine was optimized by using particle swarm optimization (pso) algorithm of punish coefficient and the kernel function parameter, building has three classification discriminant ability of support vector machine (SVM) discriminant model, finally through the credit identification ability to calculate each index index weights. The characteristic of this paper is that by combining the loss function with the support vector machine of particle swarm optimization, a new idea of credit index weight of multi-class support vector machine under particle swarm optimization is proposed, and a credit index weight system that can reflect the credit status of three types of enterprises is able to be constructed. Compared with the classical entropy weight method and logistic regression weight calculation method, the three-classification SVM weight calculation model studied in this paper has higher credit discrimination ability.

Keywords: credit evaluation; weights of indicators; the default state; triple classification support vector machine; particle swarm optimization

1 Introduction

The essence of credit index weight measurement is that the size of index weight must be able to distinguish the default status, and the greater the degree of differentiation of default status, the greater the weight. Due to fewer leasing and business services samples of small business credit and default sample amount is less, so the leasing and business services small enterprise credit assessment is a small sample and imbalance of credit evaluation, in the sample under the background of serious imbalance, the leasing and business services to small business credit evaluation is an urgent need and challenging research work.

At present, the research on credit index weights is mainly divided into the following two categories:

The first: a research on the credit index weight model which reflects the index information content

Dahooie (2021) regards Data Envelopment Analysis (DEA) as an effective objective weighting method to calculate the index weight using dynamic DEA common weighting set method [1]. On the one hand, Du Zhiping (2020) uses the super-efficient DEA method to realize the objective weighting of quantitative factors, from which it can solve the problem of low discrimination of the traditional DEA for the comprehensive efficiency value; On the other hand, the preference-dependent linear weight combination formula fuses the two weighting methods [2]. Wang Junfeng (2014) uses the Lagrange model to objectively solve the index weights so as to obtain the optimal weights for credit evaluation [3].

* Corresponding Author

The second: a research on the credit index weight model which reflects the two-category default status of default and non-default

Hu Yu-shan (2021) synthesizes different characteristics of enterprise data, determines the combination weight by using the normalized constraint with the largest variance, and establishes a credit evaluation model based on combination weighting method [4]. Chi Guotai et al. (2020) indicator weighting is to obtain two sets of indicator weights through unit redundancy discrimination degree and grey correlation on the premise of obtaining the optimal indicator combination, to establish an objective programming function with the largest amount of indicator information and the highest consistency of weighting results, and to deduce a set of optimal indicator weights [5]. Xu Zhandong (2019) uses KMV to calculate the default distance, which reflects the expected credit risk of the market. And, based on the empowerment idea that credit risk assessment should be consistent with the expected risk of the market, he builds a model to minimize the error with the expected credit risk of the market and weights the indicators [6].

There are two deficiencies in the existing credit evaluation research: First, the credit index weight model which reflects the information content of the index calculates the weight based on the degree of variation between the data, and the obtained weight lacks credit identification ability. **The second** is that although the credit index weight model of two categories of default status can distinguish the credit status of enterprises in default and non-default, it does not explain the characteristics of high default and low default of enterprises, and ignores that high default loss and low default loss are different credit characteristics.

Aiming at the above problems, this paper constructs a support vector machine tri-classification model to establish a nonlinear objective programming function, which takes the maximum dispersion of credit score as the goal and the maximum division force of index weight as the constraint, and solves the problem that the credit index weight can distinguish the default status of tri-classification.

2 Materials and Method

2.1 Difficulties of the Problem

Difficulty 1: Which method is used to calculate the credit indicator weight that is more important for the greater the ability to distinguish the multi-category default status.

Difficulty 2: Which algorithm should be used to optimize the key parameters of the credit index weight calculation model, so as to make the model more accurate and faster. Ethics approval of research.

2.2 Breakthrough of Difficulties

The solution to the first difficulty: The first stage uses the loss function to divide the real default situation into three situations: non-default, low default and high default. In the second stage, the support vector machine improvement is applied to the three classification problems to obtain the predicted values of all enterprise default cases; In the third stage, the credit identification ability of the index is obtained through calculation, and based on this, the credit index weights distinguishing various default states are obtained.

The solution to the second difficulty is to use the particle swarm optimization algorithm to optimize the key parameters of the credit index weight calculation model, search through the memory function of the particles themselves in the solution space, and finally iterate out the optimal parameters.

3 Multi-classification Support Vector Machine Model Based on Particle Swarm Optimization

3.1 Classification of Enterprise Credit Status

3.1.1 Calculation of Loss Rate

Set up: LR_i- loss rate of the I-th enterprise; L_i- the outstanding principal and interest receivable from the I-th enterprise; R_i–principal and interest receivable of the I-th enterprise. The calculation formula of enterprise loss rate is:

$$LR_i = L_i / R_i. \quad \textbf{(1)}$$

The meaning of formula (3.1): indicates the degree of default of the enterprise, the higher the principal and interest receivable, the higher the loss rate, it indicates the greater the degree of default of the enterprise, on the contrary, it indicates the smaller the degree of default.

3.1.2 The Default State Separation Parameter Selection

When classifying enterprises by using loss rate as a judgment standard, it is necessary to select an appropriate standard to distinguish between low default and high default, if the imbalance of high default and low default samples will lead to the low classification accuracy of support vector machine prediction, this paper selects 0.9 as the separation parameter through comparative analysis, and the number of high default and low default samples is relatively balanced and the degree of differentiation is the highest in different intervals where 0.9 is the critical value. Therefore, this paper selects 0.9 as the segmentation point of high default and low default status intervals, and the default status is shown in Table 1 below.

Table 1. Rules for determining default status

Serial number	LR	Default status	Number of individuals in this state
1	$LR = 0$	Non-default	87
2	$0.9 > LR > 0$	Low default	12
3	$1 > LR > 0.9$	High default	14

3.2 Support Vector Machine Model Construction

3.2.1 The Establishment of Optimization Function

Let's say: the sample set is $Z = \{(x_i, y_i)\}$, $i = 1, 2, ..., l$, where $x_i = (x_{1i}, x_{2i}, ..., x_{20i})$ represents the data set of Enterprise I on 20 classification indicators; y_i–default status of the enterprise, $y_i \in \{-1,1\}$, $y_i = 1$ indicates that enterprise I belongs to category 1, and $y_i = -1$ indicates that enterprise I belongs to category 2; l- number of samples; α_i- the langrange multiplier corresponding to each sample; $k(x_i, x_j)$- support vector machine kernel function; b-SVM classification threshold. The calculation formula of the optimal classification decision function $f(x)$ is:

$$f(x) = sign[(\sum_{i,j=1}^{l} \alpha_i y_i k(x_i, x_j)) + b]. \quad \textbf{(2)}$$

The use of formula (3.2): the result obtained from the optimal classification function is further calculated to obtain the predicted value of the default status of the enterprise and then used to measure the credit identification ability.

3.2.2 Selection of Kernel Function

Let: $k(x_i, x_j)$-support vector machine kernel function; Parameter of δ-kernel function. The kernel function formula of that support vector machine is:

$$k(x_i, x) = \exp(-\frac{\|x_i - x\|^2}{2\delta^2}). \quad \textbf{(3)}$$

Benefits of Equation (3.3): as the Gaussian radial basis function has better effect than other kernels in approximating nonlinear functions and has fewer parameters than other kernels, this paper selects the Gaussian radial basis function to construct the classification prediction model of support vector machine.

3.2.3 Support Vector Machine Parameter Optimization Based on Particle Swarm Optimization

Let's say: v-particle velocity; x-particle position; t-number of iterations; w-inertia weight; Random numbers in the r_1-[0, 1] interval; Random numbers in the r_2-[0, 1] interval; c_1- nonnegative learning factor, c_2-nonnegative social factor; p_{jd} -local optimal solution; p_{gd} - global optimal solution. The expression for the particle velocity and position update is:

$$\mathrm{v}_{jd}^{t+1} = \omega \mathrm{v}_{jd}^{t} + c_1 r_1 (p_{jd}^{t} - x_{jd}^{t}) + c_2 r_2 (p_{gd}^{t} - x_{gd}^{t}) \quad x_{jd}^{t+1} = x_{jd}^{t} + v_{jd}^{t+1} \quad j = 1, \ldots, m. \tag{4}$$

Use of Equation (3.4): support vector machine parameters are optimized through particle swarm optimization algorithm, the particles update their own speed and position according to Equation (3.4), according to the fitness of the objective function, the optimal solution is found by updating iteration in the solution space so as to find the penalty coefficient and Gaussian radial basis function parameters required by the model.

3.3 Credit Index Weight Calculation

3.3.1 The Credit Index Identification Ability

Suppose: A- default discrimination ability of all indicators (percentage of enterprises with high default, enterprises with low default and enterprises without default that are judged correctly), which can be expressed as credit discrimination ability of all selected indicators to enterprises; A_0–represents the percentage of all non-defaulting enterprises that are judged correctly; A_1–represents the percentage of all low-default firms that are judged correctly; A_2–represents the percentage of all high default firms that are judged correctly. The formula for the indicator of credit identification capability is:

$$A_j = (A_0 + A_1 + A_2)/3. \tag{5}$$

The characteristic of formula (3.5): the default state is divided into three states: high default state, low default state and non-default state, which is different from the only default state and non-default state in the past, and makes the judgment ability of credit evaluation model more powerful.

3.3.2 The Identification Ability After Deleting Item *j*

Set A_j- the default judgment ability of other indicators except the J-th indicator; A_0^j - Percentage of non-defaulting enterprises that have judged correctly for other indicators than the J-th indicator; A_1^j - represents the percentage of low-default firms that are judged correctly for indicators other than the J-th. The formula of credit identification ability after the deletion of item j is:

$$A_j = (A_0^j + A_1^j + A_2^j)/3. \tag{6}$$

Meaning of formula (3.6): credit identification ability of all remaining indicators after deleting the J-th indicator.

3.3.3 The Calculation of Credit Index Weight

Set: d_j - the degree of impact of the J-th indicator on the credit evaluation results; A_j - the credit appraisal ability of the remaining indicators except for the j indicator to all enterprises; A- the credit appraisal ability of all indicators to all enterprises. The formula for the influence of the indicator is:

$$d_j = A - A_j. \tag{7}$$

The use of Equation (3.7): reflects the degree of impact of the J-th indicator on the credit evaluation results, and is used to further calculate the weight of credit indicators.

Set: r_j - the weight of the J-th indicator; d_j–the extent to which the J-th indicator affects credit evaluation results. The credit index weight calculation formula is:

$$r_j = |d_j| / |\sum_{j=1}^{n} d_j| . \quad \textbf{(8)}$$

Meaning of formula (3.8): the weight of each indicator under the conditions established by the credit indicator system.

4 Application of Model Conclusion

4.1 Pretreatment of Enterprise Credit Indicators

4.1.1 The Construction of Index System

The credit index system in this paper contains three criterion layers, in addition, some other indicators are selected under each standard level based on the characteristics of the enterprise, the credit evaluation system in this paper contains 20 different indicators, which are derived from Literature [8, 10], and the construction of the index system is shown in columns 2 to 3 of Table 2.

Table 2. Indicator system and standardized values

(1) Serial number	(2) The standard layer	(3) The index name	(4) Indicator type	(5) Sample 1	…	(117) Sample 113
1	Inside the enterprise Financial factors	Cash recovery rate for all assets	forward direction	0.630	…	0.698
2		Ratio of net cash flow from non-current liability operations	forward direction	0.518		0.476
3		Retained Earnings Growth Rate	forward direction	0.115		0.089
4		Main business income cash ratio	forward direction	0.688		0.999
5		Shareholders' equity ratio	forward direction	0.367		0.000
6		Net profit	forward direction	1.000		0.401
7		Capital immobilization ratio	negative direction	0.519		0.568
8		Current debt ratio in earnings before interest and tax	forward direction	0.710		0.748
9	Internal non-enterprise Financial factors	Percentage of total loans recovered by enterprises through the Bank	forward direction	0.704	…	0.904
10		Brand name product level	qualitative	0.550		0.450
11		Years of working in related industries	qualitative	0.350		0.000
12		Legal representative loan default record	qualitative	1.000		0.000
13		Duration of holding the position	qualitative	0.000		0.000
14		Living conditions	qualitative	0.710		0.000
15		Enterprise credit status in recent three years	qualitative	0.500		0.000
16		Abiding by the law of the enterprise	qualitative	1.000		0.000
17		Enterprise legal disputes	qualitative	0.600		0.200
18		Mortgage score	qualitative	0.725		0.547
19	External macro-environment factors of enterprises	Industry boom index	forward direction	0.578	…	0.486
20		Engel coefficient	negative direction	0.855		0.756
21		Sample mean		0.606	…	0.366
22		Sample standard deviation		0.259	…	0.335
23	Real credit status of micro-enterprises in leasing and business services			1	…	3

4.1.2 Data Standardization

The application adopts the credit data of small enterprises in leasing and business service industries, and the used enterprise data comes from the database of a city bank in China, and the sample data span from 2019 to 2020, a total of 113 samples of enterprises are selected, among which 87 are non-default enterprises, 12 are low default enterprises and 14 are high default enterprises; The real credit default status of an enterprise is identified by numbers, where 1 indicates that the real credit status of the enterprise is non-default, 2 indicates that the real credit status of the enterprise is default, and 3 indicates that the real credit status of the enterprise is high default. The results of the true credit status of the enterprise are shown in line 21 of Table 2. By standardizing the data according to the characteristics of the indicators, the results are shown in columns 5 to 117 of Table 2.

4.2 Selection of Optimal Parameters Based on Particle Swarm Optimization

The particle swarm optimization algorithm is implemented by using matlab2018a software and libsvm3.1 toolbox, and the optimal penalty coefficient c and the optimal Gaussian radial basis function δ are obtained, the specific results are shown in Table 3.

Table 3. Support vector machine optimization parameters

Parameter	Penalty factorc	Parameter δ of Gaussian Radial Basis Function
numerical value	96.8886	0.355242

4.3 Credit Evaluation Index Weight Calculation

Step1: All indicators credit indicator identification ability

Substituting the standardized data into Equation (3.2) can obtain the predicted value of the enterprise default status, the above process is relatively complicated, support vector machine three-category prediction can be completed through the matlab program, The test set results are shown in Fig. 1, by substituting the prediction results into Equation (3.5) can calculate the credit identification ability A of all indicators to the enterprise, and the calculation result A=0.78.

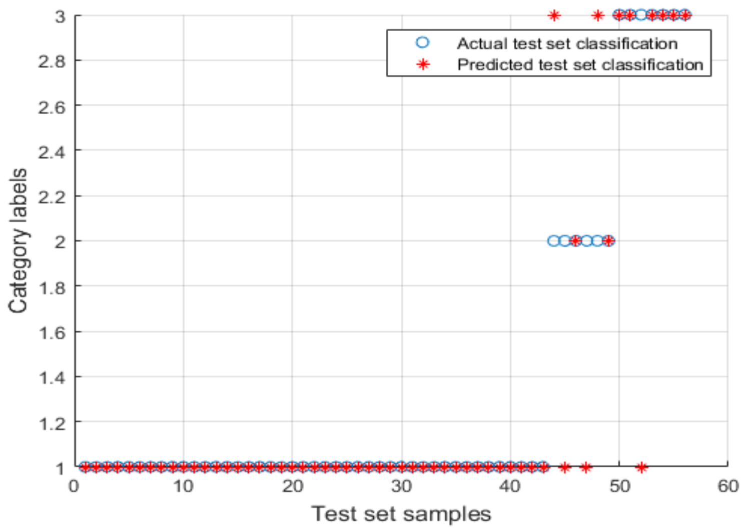

Fig. 1. Test Set Results

Step2: Identification ability after deleting item j of the index

Substituting all the index data after the deletion of item j into Equation (3.2), completing the support vector machine three-class prediction through the matlab program, and substituting the prediction result into Equation (3.6), to obtain the credit identification ability A_j of all the indexes except the deletion of item j to the enterprise. The results of A_j calculation are shown in column (4) of Table 4.

Table 4. Weights of enterprise credit evaluation indicators

(1) Serial number	(2) The standard layer	(3) The index name	(4) A_j	(5) d_j	(6) r_j	(7) Entropy weight method	(8) Logistic regression weights
1	Internal financial factors	Cash recovery rate for all assets	0.63	0.15	0.117	0.04699	0.0547
…		…	…	…	…	…	…
8		Current debt ratio in earnings before interest and tax	0.72	0.06	0.046	0.04311	0.0849
9	Internal non-financial factors	Percentage of total loans recovered by enterprises through the bank	0.73	0.05	0.039	0.00569	0.0057
…		…	…	…	…	…	…
18		Mortgage score	0.72	0.06	0.046	0.08119	0.0241
19	External macro-environmental factors	Industry boom index	0.73	0.05	0.039	0.08863	0.0184
20		Engel coefficient	0.73	0.05	0.039	0.12448	0.1398

Step3: Index influence d_j calculation

d_j is the difference between the credit identification ability of all indicators and the identification ability of all remaining indicators after the deletion of indicator j, which is expressed as the magnitude of the impact of indicator j on credit evaluation. Substituting A and A_j calculated above into Equation (3.7), the influence of each index is calculated. The calculation results of d_j are shown in column (5) of Table 4.

Step4: Calculation of Indicator Weight

In this paper, d_j is used to calculate the weight of index j, and the index influence d_j shown in column (5) of Table 4 is substituted into the credit index weight in Equation (3.8). The calculation results of r_j are shown in column (6) of Table 4.

Table 5. Comparative analysis of models

(1) The right of way	(2) Whether to reflect the credit status	(3) The default judgment accuracy
Entropy weight method	no	0
Logistic regression weights	Reflecting two classifications	89.38%
Three categories of credit index weights	Reflecting the three classifications	94.69%

4.4 Comparative Analysis of Indicator Weights

The entropy weight method and Logistic regression model are used as the comparison objects of the three-category credit index weight model, and the calculation results of index weight are shown in columns 7 to 8 of Table 4.

Through the comparison with the index weight system established by entropy weight method and binary Logistic regression model, the credit index weight system established by the multi-classification support vector machine model under particle swarm optimization has the ability to identify the three-classification default states which cannot be achieved by the two methods mentioned above, so it builds a credit index model with higher default discrimination accuracy and stronger discrimination ability, which solves the problem that small enterprises in the leasing and business service industries are in urgent need of effective credit evaluation.

4.5 Some Notes

The sample of 113 enterprises was taken as an example above, and the entropy weight method and Logistic regression model were selected to compare to illustrate the feasibility and rationality of the credit index weight system constructed in this paper. However, due to the imbalanced samples of corporate default and non-default, the sample size selected in this paper is not large enough, which is the weakness of this study. In the follow-up research, we aim to overcome the imbalance of enterprise samples, find a new method to deal with the enterprise data with unbalanced samples, and then use a sufficient number of samples to calculate the weight of credit indicators.

5 Conclusion

(1) Through the three-classification model of support vector machine, this paper obtains the weights of 20 different evaluation indexes at three different levels, represented by the cash recovery rate of all assets (internal financial factors of the enterprise), the length of service in relevant industries (internal non-financial factors of the enterprise) and Engel coefficient (external macro-environmental factors of the enterprise).

(2) Through the comparative analysis with the index weight system established by the classical entropy weight method and the logical regression calculation method, it is concluded that the three-category support vector machine weight calculation model studied in this paper has higher credit discrimination ability.

(3) The support vector machine is used to create different classifiers, which realizes the three classifications of the support vector machine, solves the three classification problems of non-default, low-default and high-default of credit status, reflects the principle that the larger the ability to distinguish the default characteristics of an index, the more important the index is, and constructs a credit index weight system with stronger discrimination and higher discrimination accuracy.

Acknowledgments

The research is supported by Inner Mongolia Natural Science Foundation of China (2020MS07009), and Science and Technology Project of Inner Mongolia Autonomous Region of China (201605053). We are grateful for the support of these organizations.

References

[1] H.-D. Jalil, H.-R.-H. Seyed, F. Shima, A novel dynamic credit risk evaluation method using data envelopment analysis with common weights and combination of multi-attribute decision-making methods, Computers & Operations Research 40(2021) 223-241.

[2] Z.-P. Du, M. Zhang, Cross-border e-commerce logistics enterprise performance evaluation, statistics and decision 36(21)(2020) 181-184.

[3] J.-F. Wang, H.-Y. Wu, B2C Enterprise Credit Evaluation Based on Improved TOPSIS Method, Soft Science 28(06)(2014) 21-24.

[4] Y.-S. Hu, C.-L. Zhu, Y. Li, Credit Evaluation Model of Road Transportation Enterprises Based on the Combination Weighting Method, Mathematical Problems in Engineering 2021(2021) 26-36.

[5] Z. Li, G.-T. Chi, Research on the credit risk of listed companies based on the maximum indicator discrimination degree and the optimal relative membership degree, China Management 2021(2021) 1-12.

[6] Y.-Q. Cheng, Z.-D. Xu, ELECTRE Ⅲ Small Business Credit Evaluation Model Based on theil index’s Revision, China Management Science 27(10)(2019) 22-33.

[7] S.-L. Yu, G.-T. Chi, X. Jiang, Credit rating system for small businesses with the greatest ability to identify default based on the indicator system and empirical evidence, China Management Science, 28(06)(2020) 38-50.

[8] Z.-J. Li, C.-R. Yang, Marco Mussetta, Establishment of the Credit Indicator System of Micro Enterprises Based on Support Vector Machine and R-Type Clustering, Mathematical Problems in Engineering 2018(2018) 1-11.

[9] M. Li, G.-J. Ma, SME credit evaluation index system construction, statistics and decision 23(2005) 61-62.

[10] Y.-Q. Cheng, Research on credit risk assessment of small enterprises based on default discrimination, Scientific research management 36(S1)(2015) 510-517.