

Research on the Forecasting of Enterprise Credit Scoring Based on SVR Model

Jiahui Yang*

School of Management, Shanghai University, Shanghai 200444, China

*Corresponding author e-mail: 18340835909@163.com

ABSTRACT. *In the competitive market environment, enterprises have to face all kinds of risks at any time. Among them, credit risk is one of the most challenging risks. As the main analysis method of credit risk assessment, how to establish and improve the credit scoring model of the company is of great significance to the enterprise. This paper proposes a support vector regression (SVR) model with kernel function to solve the forecasting problem of corporate credit scoring. In this paper, the model is applied to a group of basic information data sets of enterprises with specific credit score values, and we study the prediction effect of the model. In the experiment, linear regression model and neural network model are selected to compare with SVR model. The experimental results show that the accuracy and efficiency of SVR model are better than the other two models. Compared with the traditional expert scoring method, the SVR model can automatically predict the company's credit score according to the existing information data of the enterprise, which not only saves cost but also improves efficiency. Therefore, the model has a high reference value for dealing with the forecast of enterprise credit rating.*

KEYWORDS: *SVR, credit scoring, forecasting problem*

1. Introduction

In today's socialist market economy system, enterprises always face various types of risks at all times. Among them, credit risk is the most likely risk that causes enterprises to fall into crisis. The ability of enterprises to control and manage credit risk is directly related to the healthy, stable and lasting development of enterprises. Especially in recent years, the outbreak of the world financial crisis has caused enterprises to attach great importance to credit risk.

For an enterprise, its ability to control credit risk is directly related to the bank's decision on whether to grant loans to the enterprise and the relationship management between the enterprise and customers. It can be seen that the management and control of credit risk is closely related to the profitability of enterprises. Therefore, how to build a reasonable credit scoring system is an important topic. However,

there are few researches on the credit scoring of companies in the existing references. Scholars at home and abroad mostly focus on how to improve the personal credit scoring system. In addition, due to the single type of existing real data sets used for credit scoring, most of the data sets only have category labels. For example, the most commonly used customer information data sets for credit scoring in Germany and Australia often only give the judgment that the customer's credit status is "excellent", "good" or "bad", or give the customer's credit rating, without specific credit score value. Based on such data sets, scholars build different credit scoring models to predict the credit status of customers or companies, which is essentially regarded as a classification problem. However, only classifying the credit status of a company does not reflect the differences between enterprises in the same category. Therefore, if we can find the data set containing the value of corporate credit score, we can study the prediction of corporate credit score, which is more meaningful than classification:

(1) Helping to understand the credit difference between enterprises

Credit rating of enterprises can not well understand the credit differences between enterprises of the same level. Therefore, the use of specific credit score is conducive to understanding the credit situation of each enterprise.

(2) Saving cost and improving efficiency

Many experts are required to evaluate the credit score of a company, which not only takes a lot of time and energy, but also has low efficiency. Inversely, the SVR model can automatically forecast the credit score according to the basic data of the enterprise, which can save the cost and improve the efficiency.

2. Literature review

The first credit scoring model was designed by Altman. The most common traditional models for credit scoring are linear discriminant analysis and logistic regression. The weakness of linear discriminant analysis is that it assumes that there is a linear relationship between variables, but the variables in the real data sets are usually nonlinear, and it does not necessarily conform to the assumption that the error is subject to multivariate normal distribution. The essence of logistic regression is still linear regression, which can be used to solve the two classification problem. That is, the prediction value is 0 or 1. However, it does not need the assumption that the error is subject to normal distribution. The above two methods are based on the assumption that there is a linear relationship between variables, so the prediction effect is lack of accuracy.

Chinese scholars Deng Yunsheng and Liu Liya (2004) made a detailed analysis and discussion on the two main rating methods (expert evaluation method and model method) adopted by banks. For expert evaluation method, the difficulty of using 5C method or other expert evaluation methods lies in how to set the optimal weight. While expert evaluation is subjective, and the weight set by each expert may be different. For the common models of foreign banks, the core is to establish PD and

LGD risk factor models. This paper analyzes the advantages and disadvantages of these two methods in detail, and puts forward suggestions on the internal credit rating of Chinese banks. Tan Qingmei, Wu Jinke and Zhao Liming (2009) established the enterprise credit rating model based on BP neural network. Through the selection of hidden layer nodes, training functions and output functions of neural network, the credit rating prediction model was established. The experimental results show that when the number of hidden layer nodes of trainlm is 17, the neural network has a good fitting effect. Pang Sulin, He Yizhou et al. (2017) established a three-level credit scoring model including enterprises, industries and regions for four different enterprise hierarchies, and gave a credit scoring formula. This scoring model is applied to the relevant data sets of the company's credit loan application. The experimental results not only give the credit rating status of the company and its subsidiaries in different hierarchical structures, that is, "excellent" or "good", but also give the credit score numerical conditions of the loan application: only the enterprises with a score of 27 can apply for the loan.

In addition to the methods mentioned above, support vector machine (SVM) is the most commonly used method in the study of credit scoring. It has been proved to be one of the techniques that can produce ideal classification effect. Baesens et al, a foreign scholar, have studied 17 different real credit data sets based on different classification techniques. They use SVM with kernel function and LS-SVM to compare with other methods, and use grid optimization method to adjust parameters. Experimental results show that SVM ranks the highest in performance. Gestel et al. used LS-SVM model to rate banks and compared the results with linear regression model and MLP. The results show that the accuracy of LS-SVM classifier is better than other methods. Zhang Mu and Zhou Zongfang (2010), domestic scholars, used SVM model in the study of corporate credit rating, combined with real code accelerated genetic algorithm to calculate the weight of each attribute. Experimental results show that SVM model performs better in classification accuracy.

In view of the SVM model can produce better classification effect in the classification of credit score, the application of support vector regression (SVR) model to the prediction of credit score will also have better prediction accuracy. Due to the limitations of existing data types, there are few scholars at home and abroad who use SVR model to study the problem of forecasting corporate credit scores. Therefore, the SVR model used in this paper has a good reference value for solving this kind of problem.

3. Main research theories

3.1 Linear regression model

The purpose of univariate linear regression is to find a set of parameters: (β_0, β_1) that satisfy: $y = \beta_0 + \beta_1 x + \varepsilon$, the minimum sum of the vertical distance from all points on the line to the actual point. The objective function can be expressed as follows:

$$(\beta_0, \beta_1) = \underset{\beta_0, \beta_1}{arg\ min} L(\beta_0, \beta_1) \quad (1)$$

$L(\beta_0, \beta_1) = \sum (y^i - \beta_0 - \beta_1 x^i)^2$ is the sum of the squares of the vertical distances, and the partial derivatives are obtained as follows :

$$\beta_1 = \frac{\sum (x^i - \bar{x})(y^i - \bar{y})}{\sum (x^i - \bar{x})^2} \quad (2)$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x} \quad (3)$$

The vertical error between the actual value and the predicted value is expressed as follows:

$$\varepsilon_i = y^i - \hat{y}^i = y^i - \beta_0 - \beta_1 x^i \quad (4)$$

The linear regression model is based on the following assumptions:

- (1) There exists a linear relationship between variables.
- (2) Follows normal distribution.
- (3) The error variables are independent and have no correlation.
- (4) Reject extrapolation.

For multiple linear regression model, its principle is the same as that of single linear regression model. However, when doing multiple linear regression, if the number of variables is large, there may appear multiple collinearity, which will have a negative impact on the fitting effect.

3.2 Neural network model

The neural network model consists of input layer, hidden layer and output layer. For two-layer neural network, its training process consists of the following four steps:

- (1) Initialization weight.
- (2) Input tuple from the input of neural network.
- (3) Calculate the linear sum of the input value and the output value according to the activation function.
- (4) Calculate the deviation and modify the weight according to the deviation.

Every sample in the training set is trained in order to make the actual output value approach the ideal output value.

The two-layer neural network is suitable for solving linear problems, and its convergence speed is fast. The multi-layer neural network, which has a wider application range, is suitable for solving multi-element nonlinear problems, but the hidden layer error can not be calculated directly. In neural network, the number of hidden layers is determined according to specific problems. In general, the more

complex the problem is, the more hidden layers it needs, the more computation it takes, and the worse the generalization ability of the network. In addition to the number of hidden layers, the selection of activation function and output function, the number of hidden layer nodes and the selection of training function will affect the prediction performance of neural network. Neural network may produce local minimum instead of global minimum, and it is prone to over fitting. Therefore, we often use support vector machines(SVM) instead of neural network to study credit score classification.

3.3 Support vector regression (SVR) model with kernel function

Suppose a set of training data set: $T = \{(x^i, y^i), i = 1, \dots, n\}$ is given, $x^i = (x_1^i, x_2^i, \dots, x_m^i) \in R^m$ represents the i th training point in m Dimensional space, $y^i \in R$ is the corresponding output function. SVR model can be summarized as follows:

$$\begin{aligned} \min \quad & \frac{1}{2} \|W\|^2 + \frac{C}{n} \sum_{i=1}^n (\xi_i + \xi_i^*) \\ \text{s.t.} \quad & y^i - f(x^i) \leq \varepsilon + \xi_i \\ & f(x^i) - y^i \leq \varepsilon + \xi_i^* \\ & \xi_i, \xi_i^* \geq 0 \end{aligned} \tag{5}$$

Where $f(x^i) = W \times \varphi(x^i) + b$, $W \in R^m$ is the weight vector, and $\varphi(x)$ denotes a non-linear function of mapping inputting training data to kernel space, loss function $\varepsilon > 0$, C represents penalty parameter, ξ_i and ξ_i^* are different slack variables.

The above model can be understood as follows: the absolute value of the deviation between the actual output value y^i and the predicted value $f(x^i)$, and the deviation within this range can be tolerated. For the point whose deviation is beyond the range, we introduce slack variable ξ_i and ξ_i^* to establish soft boundary.

In order to simplify the calculation, the convex quadratic optimization problem is transformed into lagrange function:

$$\begin{aligned} L_{(w,b,\xi^i,\alpha^i)} = & \frac{1}{2} \|W\|_2^2 + \frac{C}{n} \sum_{i=1}^n (\xi_i + \xi_i^*) - \sum_{i=1}^n (\xi_i \eta_i + \xi_i^* \eta_i^*) - \sum_{i=1}^n \alpha_i (\xi_i + \varepsilon + y^i - W \times \varphi(x^i) - b) \\ & - \sum_{i=1}^n \alpha_i^* (\xi_i + \varepsilon - y^i + W \times \varphi(x^i) + b) \end{aligned} \tag{6}$$

$\alpha_i, \alpha_i^*, \eta_i, \eta_i^*$ are lagrange multipliers. By minimizing W, b, ξ, ξ^* , the above problem can be transformed in the following way:

$$\begin{aligned}
 \max \quad & \left(\sum_{i=1}^n (\alpha_i - \alpha_i^*) y^i - \varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) - \frac{1}{2} \sum_{i=1, j=1}^n (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) K(x_i, x_j) \right) \\
 \text{s.t.} \quad & \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0 \\
 & 0 \leq \alpha_i, \alpha_i^* \leq \frac{C}{n}
 \end{aligned} \tag{7}$$

Because $\varphi(\mathbf{x})$ is an unknown high-dimensional function, the kernel function $K(\cdot, \cdot)$ satisfying Mercer condition is introduced to replace the inner product function. The common kernel functions are gaussian kernel function, quadratic kernel function and polynomial kernel function.

When the optimal solution of the model is $\bar{\alpha} = (\bar{\alpha}_1, \bar{\alpha}_1^*, \dots, \bar{\alpha}_n, \bar{\alpha}_n^*)^T$, the regression function can be expressed as follows:

$$f(x) = \sum_{i=1}^n (\bar{\alpha}_i^* - \bar{\alpha}_i) K(x^i - x) + \bar{b} \tag{8}$$

Compared with the above two methods, the accuracy of SVR model with kernel function is obviously improved, and the operation efficiency is also faster.

4. Experiments

4.1 Real data set

In order to verify the validity of the model, we selected the basic information data of 100 enterprises, including the number of cooperation with other enterprises within two years, the number of delivery delays within two years, the number of credit defaults in two years, and the number of substandard goods. The basic information is described in table 1:

Table 1 Basic information description statistics of 100 enterprises

	Cooperation	Delivery delays	Credit defaults	Substandard goods	Credit score
mean	129.2547	0.0189	0.066	0.5566	2.8437
max	772	1	3	5	3.8571
min	0	0	0	0	1.1667
std	160.7703	0.1361	0.3439	0.9425	0.6828

4.2 Data pre-processing

We use the normalization pre-processing method to normalize the original data set to an interval [0,1]. The formula is as follows:

$$x_j^i = \frac{x_j^i - \min_{k \in \{1, \dots, n\}} x_j^k}{\max_{k \in \{1, \dots, n\}} x_j^k - \min_{k \in \{1, \dots, n\}} x_j^k}, \quad i = 1, \dots, n \quad j = 1, \dots, m \tag{9}$$

Where the numerator represents the difference between the attribute value of column j and the minimum value in the attribute of the column, and the denominator represents the difference between the maximum value and the minimum value in the attribute of the column.

4.3 Selection of relevant parameters of the model

All experiments are carried out under Matlab 2018a version. Before formal experiments, we need to select the relevant parameters of three methods involved in this paper. We select 80% of the data as training samples, and randomly generate 20 sets of training data sets. Next, we use the value of MAPE (mean absolute error) to measure the prediction accuracy of the model. The specific expression is as follows:

$$\text{mape} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_{\text{predict}} - y_{\text{real}}}{y_{\text{real}}} \right| \quad (10)$$

(1) For the SVR model with kernel function, the choice of parameters has a great influence on the prediction accuracy of the model, especially the choice of penalty parameter C . In this experiment, the g value of Gaussian kernel function and polynomial kernel function denotes the default value. We set the range of C between 17 values. Each experiment uses five cross methods to adjust parameters. Then, we calculate the MAPE mean value of different C values, and choose the C value corresponding to the minimum MAPE mean value as the best C value of the kernel function. Finally, the best C value of Gaussian kernel function is 8, the best C value of linear kernel function is 0.25, and the best C value of polynomial kernel function is 256. We fix the C value of each kernel function, train each training data set again, and keep the corresponding model.

(2) In the neural network model, because the experimental data set is small, we set the hidden layer number as a single layer, the training function uses the default value 'trainlm', the activation function is set to 'logsig', the output function is set to 'purelin', the number of hidden layer nodes is set to 10, 15 and 20 respectively, and the experiment is divided into three groups. Each experiment keeps the network model of training. The results show that when the number of nodes is 20, the mean value of MAPE is the smallest.

4.4 Experimental result and analysis

According to the 20 training datasets in the above pre-experiment, we get the corresponding test datasets. The experimental results are as follows:

Table 2 experimental results

Model	mape				cputime
	mean	var	max	min	
SVR_GausKernel	0.1378	0.0017	0.2706	0.0691	0.0654
SVR_LinearKernel	0.1407	0.0011	0.2367	0.0975	0.0602
SVR_PolyKernel	0.1690	0.0013	0.2671	0.1191	0.0889
linear regression	0.1654	0.0028	0.3143	0.0901	0.0578
neural network	0.1589	0.0031	0.3382	0.0795	2.3250

(1) The MAPE mean of SVR model with Gaussian kernel function is the smallest of the five methods. The MAPE mean value of SVR model with polynomial kernel function is the largest of the five methods, followed by linear regression model. This suggests that their predictions are not ideal. The MAPE variance of SVR model with linear kernel function is the smallest and the model is the most stable.

(2) The variance of MAPE of linear regression and neural network model is larger than the former three models. This indicates that both two comparison methods are still unstable. In addition, the neural network model also has the longest running time.

In terms of prediction accuracy, model stability and running time, SVR model is obviously superior to the other two models. Next, we have carefully analyzed several reasons for the above experimental results:

(1) As the most stable linear kernel function, SVR has fewer parameters and only needs to adjust the C value. In the process of pre-experiment parameter adjustment, the frequency of parameter C equals to 0.25 is 0.9, which means that the C value used in almost every training data set is the optimal C value for this training set. Therefore, each experiment is predicted by a model trained with the same optimal parameters, and the effect is naturally stable.

(2) SVR with polynomial kernel function belongs to global kernel function, and it can also affect the value of polynomial function for the far point. In addition to penalty parameter C, the choice of times d will also have an impact on the prediction effect. If the value of D is too large, the spatial dimension of its mapping will also be large, which is prone to over fitting. Gauss kernel function is different from polynomial kernel function in that it belongs to local kernel function and is not easily affected by outliers. So in the experiment, the mean value of MAPE of SVR model with polynomial kernel function is the highest, and the prediction accuracy is the lowest.

(3) For linear regression model, it is easy to be affected by outliers, because when it uses the least square estimation principle, the default weight of the square of the vertical error of all terms is 1. Once outliers exist in the data, the fitting line will deviate. In addition, because of the characteristic of "refusing extrapolation" of linear regression model, it is easy to be limited by the value range of independent variables. The larger the value range, the worse the fitting effect.

(4) Neural network has a good prediction effect on the training data samples, but the prediction effect is not ideal for the test set samples in the formal experiment. This also proves that the fitting effect of the data points in the sample is good, but for the data points outside the sample, its fitting effect is poor, and the neural network has a long running time and low efficiency. So in the actual prediction problem, the SVR model is often chosen instead of the neural network model.

5. Conclusion

At present, China has not formed a set of mature system architecture for the construction of corporate credit scoring system, and the existing research direction is mostly focused on the construction of personal credit scoring system. In addition, due to the limitation of data types, credit scoring problems are often regarded as classification problems. Traditional corporate credit scoring needs a lot of time and effort from experts, but SVR model can automatically predict corporate credit score according to the basic information data of the company, so as to understand the credit difference between different enterprises. This is more meaningful for enterprises, banks and other financial institutions.

In the next research work, we consider adding robust regression model as a comparison method. Compared with linear regression model, robust regression is not easily affected by outliers. In addition, because SVR model contains kernel function, we need to adjust parameters to get better prediction effect, that is to say, the setting of parameters has a great impact on the prediction effect. In recent years, many scholars have been devoted to the research of SVM model without kernel function for disease monitoring, credit scoring and other classification problems. Therefore, we will consider further improving the traditional SVR model in the future, so that the SVR model has a higher reference value and practical significance.

References

- [1] Deng Yunsheng, Liu Liya. A comparative study of internal credit rating methods of commercial banks [J]. Contemporary finance and economics, 2009, Issue 9.
- [2] Tan Qingmei, Wu Jinke, Zhao liming. Research on credit evaluation of small and medium-sized enterprises based on BP neural network [J]. Journal of northwest agricultural and Forestry University of science and Technology, 2009, vol.9, no.5.
- [3] Pang Sulin, He Yizhou, Wang Shouyang, Jiang Hai. Multi level cross credit scoring model and application based on risk environment [J]. Journal of management science, 2017, 20 (10): 57-69.
- [4] Zhang Mu. Enterprise credit evaluation model based on multi-objective programming and support vector machine [C]. The third prize winning paper collection of Guizhou natural science excellent academic paper, 2010, vol.9, no.10.

- [5] Luo ShuTing, Cheng BoWen, Hsieh ChunHung. Prediction model building with clustering-launched classification and support vector machines in credit scoring [J]. Expert Systems With Applications, 2009, 36 (4).
- [6] Weimin Chen, Chaoqun Ma, Lin Ma. Mining the customer credit using hybrid SVM technique [J]. Expert Systems With Applications, 2009, 36 (4).
- [7] Zhou Ligang, Lai KinKeung, Yu Lean. Least squares support vector machines ensemble models for credit scoring [J]. Expert Systems With Applications, 2009, 37 (1).
- [8] Terry Harris. Credit scoring using the clustered support vector machine [J]. Expert Systems With Applications, 2015, 42 (2).
- [9] Yao Ping, Lu Yongheng. Neighborhood rough set and SVM based hybrid credit scoring classifier [J]. Expert Systems With Applications, 2011, 38 (9).
- [10] Chi BoWen, Hsu ChiunChieh. A hybrid approach to integrate genetic algorithm into dual scoring model in enhancing the performance of credit scoring model [J]. Expert Systems With Applications, 2012, 39 (3).
- [11] Tang Bo, Xia Min. A New Fuzzy Support Vector Machine for Credit Scoring [J]. Applied Mechanics and Materials, 2012, 1487 (109).
- [12] S. Harikrishna, M.A.H. Farquad, Shabana. Credit Scoring Using Support Vector Machine: A Comparative Analysis [J]. Advanced Materials Research, 2012, 1566 (433).
- [13] Zhou LiLai, K.K. and Yen, J. Credit scoring models with AUC maximization based on weighted SVM [J]. International Journal of Information Technology and Decision Making, 2009, 4, 677-696.
- [14] Wang, Y., Wang, S. Lai, K. A new fuzzy support vector machine to evaluate credit risk. IEEE Transactions on Fuzzy Systems, 2005, 13, 820-831.