ISSN 2616-5902 Vol. 3, Issue 5: 127-132, DOI: 10.25236/AJBM.2021.030520

Research on Early Warning Model of Financial Crisis Based on SCAD-SVM

Yuan Yan[1], Ling Zhao[2]*, Yutu Tang [3], Tianxiao Luo[4]

[1]Graduate School, Jose Rizal University, Manila 0900, Philippines
[2]Business School, Changsha Commerce and Tourism College, Changsha, Hunan 410000, China
[3]School of Geography and tourism, Hunan University of Arts and Science, Changde, Hunan 415000, China
[4]Guangxi Forest Resources and Environment Monitoring Center, Nanning, Guangxi 530000, China

Abstract: *After more than half a century of development, there are numerous empirical analyses on financial early warning models, most of which take listed companies as the research object. The financial crisis early warning of listed companies is a small sample, and the application of neural network and logistic regression model has the problem of over-fitting, which leads to the effect of the model is not obvious. In recent years, the support vector machine method given risk minimization criteria has been widely used in financial crisis early warning, but few scholars combine SVM and penalty letter SCAD methods for financial early warning analysis. Therefore, this paper proposes an improved variable selection method Scad and support vector machine (SVM) combined model algorithm to select the best indicators and used for financial crisis early warning, and then based on the Shanghai and Shenzhen A-share listed companies as the research sample comparison Empirical effects of SVM, Lasso-SVM and Dantzig-SVM, the research results show that the selection of variables can greatly improve the accuracy of the financial crisis early warning model, and the SCAD method in the variable selection method has more advantages than Lasso and Dantzig. The SCAD-SVM model proposed has an accuracy rate of over 96% on both the training set and the test set in this paper. The model has a good classification effect and a strong economic interpretation ability. The research results of this paper not only enrich the financial early warning model, but also greatly improve the accuracy of financial crisis early warning, and provide a theoretical basis for the scientific decision-making of enterprises and shareholders.*

Keywords: *Support vector machines; Financial crisis warning; Smoothly clipped absolute deviation*

1. Introduction

With the interaction of globalization and the world Internet, the business of a company can be developed globally. If a company has a financial crisis, it means that the scope of the crisis caused by the crisis is getting larger and larger. Therefore, the establishment of an effective financial crisis early warning model becomes More and more important, the purpose is to be able to timely, accurately and quickly carry out early warning, monitoring and resolution of the financial crisis of the enterprise, which has quite important theoretical value and practical application value.

The financial crisis models that have been proposed so far include univariate models, multivariate models, conditional probability models, and artificial intelligence models. In particular, the logistic regression model in artificial intelligence has appeared in the literature since its birth, and its predictive ability is also constantly enhanced. However, through practical application, it is found that the model also has its shortcomings. Logistic regression cannot be used to solve the nonlinear problem. Because the Logistic decision surface is linear, it is difficult to deal with the problem of data imbalance. In addition, the method itself cannot filter variables, and must use other methods for variable selection.

Artificial neural networks are also very popular in the 21st century because they have special advantages, such as handling noise: after an artificial neural network is trained, even if part of the input data is missing, it still has the ability to identify samples; and there are other advantages is not easy to damage: because the artificial neural network represents data in a distributed way, it can still work normally when some units are damaged,. However, according to general reports, the artificial neural network model requires A large amount of training data is used to estimate the distribution of input patterns, and due to its overfitting nature, it is difficult to generalize the results. In addition, it also

ISSN 2616-5902 Vol. 3, Issue 5: 127-132, DOI: 10.25236/AJBM.2021.030520

depends entirely on the experience or knowledge of the researchers to preprocess the data to select control parameters including relevant input variables, hidden layer size, learning rate and momentum(Lawrence, Giles, & Tsoi, 1997; Min & Lee, 2005).

Support Vector Machine (SVM) is a generalized linear classifier (generalized linear classifier) that performs binary classification of data according to supervised learning (supervised learning). Its decision boundary is the maximum margin for solving the learning sample. Hyperplane (maximum-margin hyperplane) .The SVM developed by Vapnik (1995,1998) has many attractive features and excellent generalization performance, so it is popular on many issues. In addition, the principle of structural analysis minimization (SRM) has been well embodied in SVM(V. Vapnik & Vapnik, 1995; V. N. Vapnik, 1998). It is based on SRM and achieves the greatest generalization ability by minimizing the upper limit of the promotion error risk. This principle has been proved to be better than the traditional Neural networks use the principle of minimization (ERM) to have advantages. What SRM obtains is to minimize the upper limit of generalization error, while what ERM obtains is to minimize the error of training data. Therefore, it can be obtained that the solution of SVM may be globally optimal, while other neural network models often fall into local areas optimal solution, and there will be overfitting, and there will be no overfitting in SVM (Hearst, Dumais, Osman, Platt, & Scholkopf, 1998; Kim, 2003).In the example, Fan.A.etc first used SVM to establish a financial crisis warning model; Cristianini & Shawetaylor used table search technology to optimize the parameters of the SVM kernel function to establish an SVM model(Cristianini & Shawetaylor, 2000). The prediction effect of the model is optimized for MDA, LR and ANN; Some scholars have proposed LS-SVM, ODR-ADASYN-SVM, GS-SVM, GWO-SVM, etc. Based on SVM, improved combination model methods to predict corporate financial crisis, which the effect is also very good, but few scholars combined the penalty functions SCAD and SVM are used to predict the financial crisis of enterprises. Based on , this paper intends to use the improved model SCAD-SVM to warn the financial crisis of listed companies in China in order to obtain a better prediction effect.

2. Theoretical background

2.1 Smoothly Clipped Absolute Deviation method

SCAD (Smoothly Clipped Absolute Deviation) is an improvement of the LASSO method in the penalty function. This method can not only estimate the corresponding coefficients, but also effectively select significant model variables. More importantly, FAN and PENG mentioned when $p = o(n^{\frac{1}{3}})$, SCAD has the oracle property, which is a property that the Lasso function does not have(Fan & Peng, 2004). It can process data with singular matrices to generate sparse solutions. At the same time, it can optimize the likelihood function with penalty terms, and it is more stable than the LASSO method. Well, it is easier to implement in terms of calculations. The principle of SCAD screening variables is to retain the larger parameters in the initial model, and compress the coefficients of the variables with smaller parameters to zero, which can effectively reduce the model deviation, and is superior to the LASSO method in terms of theoretical properties and applications. In certain mathematical conditions, the penalty function of the SCAD method can be given, and its derivative is defined as follows:

$$P'_{\lambda}(\theta) = \lambda \left\{ I(\theta \le \lambda) + \frac{(a\lambda - \theta)_{+}}{(a-1)\lambda} \quad I(\theta > \lambda) \right\} \tag{1}$$

Among them, a>2 and λ >0, the estimate obtained under this penalty function has good theoretical properties, the most important is the oracle property, if you choose the appropriate adjustment parameters, then the obtained estimate is the estimate under the real model, usually a = 3. 7.

At present, the SCAD penalty variable screening method has achieved a series of results, which are mainly used in generalized linear models, parameter analysis models and linear regression models. In 2010, Jianfeng Zhang proposed to simulate the influence of SCAD on the threshold parameter adjustment coefficient on the model under the BIC and GCV criteria. The research results show that BIC is better than GCV; (Jianfeng Zhang, 2010); Jing Cui based on the linear model penalty likelihood study, the method is applied to the generalized linear model, which proves that SCAD has the oracle property; in 2012(Jing Cui , Peng-jiang Guo, &Zhiming Xia, 2011); 2020 Xiaomin Luo will SCAD is

ISSN 2616-5902 Vol. 3, Issue 5: 127-132, DOI: 10.25236/AJBM.2021.030520

applied to the research of least absolute regression problem, and two good proofs have been obtained, which prove that the original problem and the relaxation problem have the same global optimal solution and optimal value under certain conditions(Xiaomin Luo & Dingtao Peng, 2020). Although the SCAD method has many excellent properties, Its theoretical properties and practical applications need to be further studied. This article applies the SCAD method to the SVM model to further study their properties and applications.

2.2 SCAD-SVM method

Combining with many scholars who used LASSO and SVM to analyze and optimize the model, this article discusses the application of the practical SCAD variable screening method in the high-dimensional function in the SVM model, and uses examples to practice the model method. The SCAD function is symmetric and non-convex. Although it has the same form as the L1 penalty when it is near 0, the expression of the SCAD penalty is as follows:

$$p_{\lambda}(|\omega|)=\begin{cases}\lambda|\omega| & |\omega|\le\lambda\\ \dfrac{\left(|\omega|^2-2a\lambda|\omega|+\lambda^2\right)}{2(a-1)} & \lambda<|\omega|<a\lambda\\ \dfrac{(a+1)\lambda^2}{2} & |\omega|>a\lambda\end{cases} \tag{2}$$

Among $a>2,\lambda>0$ are the training parameters. We expect the function to preserve continuity, so the form of SCAD-SVM here is as follows:

$$\min_{b,\omega}\frac{1}{n}\sum_{n=1}^{n}\left[1-y_i\left(b+\omega\cdot h(x_i)\right)\right]+\sum_{j=1}^{q}p_{\lambda}\left(|\omega_j|\right) \tag{3}$$

In the formula, the objective function maintains the nature of the hinge function (Hinge loss $[1-yf(x)]$ function is a convex function) and the SCAD penalty, and the parameters λ play a role in weighing the relationship between data fitting and complexity.

3. Empirical analysis

3.1 Data collection and Pretreatment

The database used in this study was obtained from the Shanghai Stock Exchange and Shenzhen Stock Exchange A-share listed companies. The selected companies are normal listed companies in 2019 and listed companies that have been specially dealt with. If the financial crisis warning of listed companies is to be carried out, the basis is whether the company has been approved by the China Securities Regulatory Commission. It will be dealt with specially, and the basis in foreign countries is whether the company is bankrupt. Due to China's special national conditions, very few listed companies will usually declare bankruptcy, and generally choose to restructure or merge. According to the annual report disclosure system of listed companies in China, the time of special treatment belongs to the same year as the financial report of the company in the previous year. Therefore, it is not much reference value for companies to use the data of the previous year to predict the crisis, and it will also exaggerate the model ability to predict.

As Ohlson pointed out, using the information obtained after bankruptcy to predict bankruptcy will overestimate the predictive power of the bankruptcy model. Therefore, summing up the experience of the predecessors, combined with the actual national conditions of China, we usually choose the data of the first third year of the specially processed company to predict the change trend of the company's financial status. If 2019 is the year T as an example, the best prediction sample is the data of the year T-3, that is, the data of 2016 as the prediction. In view of the collectability and validity of financial indicator data, after excluding companies with too many missing data in the sample, this article finally selected 215 ST companies and 1099 normal companies that were specially processed in 2019, and a total of 1,314 valid samples were collected. Collected relevant data in the year T-3 (2016) as research

ISSN 2616-5902 Vol. 3, Issue 5: 127-132, DOI: 10.25236/AJBM.2021.030520

data, and used Python 3.7 for data processing. Most of the data in the sample came from the CSMAR database, and the insufficient data came from the annual report data on the EAST MONEY website.

Obtaining experimental data from the website found that different evaluation indicators have different dimensions and dimensional units. Considering that the dimension inconsistency between different indicators will cause unnecessary errors to the model, we use normalization to preprocess the data. In order to solve the comparability between the data indicators, get the standard data value of the normal distribution, and standardize each financial indicator. The formula is as follows:

$$x_i' = \frac{x_i - \min_i}{\max_i - \min_i} \tag{4}$$

Among $\min_i$ is the smallest value in the data set, and $\max_i$ is the largest value in the data set. Through normalization deviation, the original data is linearly transformed, so that the result value is mapped to [0,1], thereby solving the dimension problem.

3.2 Index selection

Two important contents of the financial crisis early warning model are the selection of the model and the construction of early warning indicators. Model is the application of predictive algorithms, indicators are the in-depth mining of financial crisis early warning, both of them affect the accuracy of listed companies' early warning. Therefore, the effect of the crisis model depends on the generalization ability of the model algorithm and the performance of the model indicators selection, both are indispensable in the research.

The dependent variable in this article is the company "ST" (indicated by Y=1) in the A-share market as a sign of financial crisis. The independent variable indicators come from the five aspects of the company's solvency, operating capacity, profitability, cash capacity and development capacity, providing a more comprehensive financial indicator for alternatives (shown in Table 1).

Table 1 List of financial indicators

Indicator type	Financial ratio
Solvency	χ_1 Current ratio χ_2 Quick ratio; χ_3 Cash ratio; χ_4 Working capital; χ_5 Assets and liabilities; χ_6 Equity ratio χ_7 Net cash flow from operating activities/total liabilities
Management capacity	χ_8 Accounts Receivable Turnover Rate; χ_9 Inventory turnover; χ_{10} Business cycle; χ_{11} Accounts payable turnover rate; χ_{12} Working capital (capital) turnover rate; χ_{13} Cash and cash equivalent turnover rate; χ_{14} Liquid assets turnover rate; χ_{15} Turnover rate of fixed assets; χ_{16} Capital intensity
Profitability	χ_{17} Return on assets; χ_{18} Long-term return on capital; χ_{19} Net profit margin (ROA); χ_{20} Net profit margin of current assets; χ_{21} Net profit margin of fixed assets; χ_{22} Return on Equity; χ_{23} return on investment; χ_{24} EBIT; χ_{25} Cash to total profit ratio; χ_{26} Return on invested capital
Cash capacity	χ_{27} Net cash content of net profit; χ_{28} Net cash content of operating income; χ_{29} Net cash content of operating profit; χ_{30} Depreciation and amortization; χ_{31} Free cash flow of equity; χ_{32} Total cash recovery rate; χ_{33} Cash meet investment ratio; χ_{34} Operating Index; χ_{35} Cash fit ratio; χ_{36} Cash reinvestment ratio
Development ability	χ_{37} Capital preservation and appreciation rate; χ_{38} Capital accumulation rate; χ_{39} Growth rate of fixed assets; χ_{40} ROE growth rate; χ_{41} Basic earnings per share growth rate; χ_{42} Net profit growth rate; χ_{43} Operating profit growth rate; χ_{44} Comprehensive income growth rate χ_{45} Total operating income growth rate; χ_{46} Sustainable growth rate; χ_{47} Growth rate of net assets per share

ISSN 2616-5902 Vol. 3, Issue 5: 127-132, DOI: 10.25236/AJBM.2021.030520

3.3 Experimental results and analysis

This article uses PYTHON3.7 for simulation experiments. In the experiment, the ST Company and normal company in 2016 are divided into 80% training set and 20% test set according to the principle of random allocation. In order to evaluate the effectiveness of the proposed model SCAD-SVM, we conducted a comparative analysis from two aspects. First, this paper chose three variable screening methods (respectively SCAD, Lasso, Dantzig) combined with SVM to compare and analyze which method can be used to improve the prediction accuracy of the model; secondly, when the variable selection method and feature set are determined, the variables are not selected, and the original support vector machine model is used to calculate the type I error rate, type II error rate and accuracy rate of the sample , Through the comparison of the results to filter the variables whether to improve the accuracy of the model.

There are three types of output results of each model, Type 1 error rate, Type II error rate and overall prediction accuracy. The so-called type I error rate refers to the probability of misjudged ST company as a normal company, the type II error rate is the probability of misjudge a normal company as ST company, and the overall prediction accuracy rate is the average value of the model's ten training sessions. Using 2016 company financial data as the training sample, bring it into the combined model, first filter out the number of feature variables, and then obtain the first and second error rates of the training set and the training set accuracy. The results are shown in Table 2.

From the results in the table, it can be seen that the accuracy of the original model without variable screening is low. Among the three combination models, the accuracy of SCAD-SVM and Dantzig-SVM is very high, indicating that under these two combinations, The screening results of feature variables are particularly good, and the selected values are particularly suitable for the two-class classification of the model, but the accuracy of the other combination Lasso-SVM is only 4% higher than SVM, indicating that the Lasso method does not perform well in this model. (See Table 2).

Table 2 Table of classification results in the training set (unit: %)

T-3 model	I error rate	II error rate	ACC
SVM	16.84	0	83.64
SCAD—SVM	2.09	0.19	97.11
Lasso—SVM	12.84	0.57	87.54
Dantzig—SVM	4.09	0.38	96.65

Whether the model is built well does not lie in the higher the accuracy rate on the training set, but whether the model still maintains a high accuracy rate in the case of new samples, that is, the accuracy rate on the test set. If it is very high on the training set, but not ideal on the test set, perhaps the model has sample imbalance or overfitting. See Table 3 for the accuracy list of the model on the test set. It can be seen from the table that the accuracy of a single model is not much different from that on the training set, indicating that there is no imbalance in the model samples; the Scad-SVM model is only better than the Dantzig-SVM model is 2.56% higher, indicating that the sample size does not have over-fitting, the sample data is processed very well, and the accuracy rate is very high; Lasso-SVM has a mediocre performance on the test set, which is always more than ten percentage points worse than Scad-SVM. All in all, the accuracy of a single model is not as high as that of the combined model. The combined model uses the SCAD penalty function screening variable method and the support vector machine to establish a financial early warning model that has good discriminant ability and strong adaptability.

Table 3 Table of classification results in the test set (unit: %)

T-3 model	I error rate	II error rate	ACC
SVM	14.45	0	82.24
SCAD—SVM	1.52	0.76	96.86
Lasso—SVM	12.93	0.38	87.19
Dantzig—SVM	3.04	0.38	94.3

ISSN 2616-5902 Vol. 3, Issue 5: 127-132, DOI: 10.25236/AJBM.2021.030520

4. Conclusions

(1) Based on the analysis of various existing financial crisis early warning models, this paper proposes an improved variable selection method Scad combined with support vector machine (SVM) model algorithm.

(2) The variable selection method can greatly improve the accuracy of the financial crisis early warning model, and the SCAD method in the variable selection method has more advantages than Lasso and Dantzig.

(3) The accuracy of the SCAD-SVM model proposed in this paper on the training set and the test set is as high as 96% or more. The model has a good classification effect and a strong economic interpretation ability.

5. Discussion

The model not only has the function of automatically selecting variables, but the sparsity of the model is also reflected. More importantly, the model has a high accuracy rate on the test set, indicating that the model has a wide range of adaptability in new samples to consider applying the SCAD-SVM model to more different types of sample companies. At the same time, we can consider adding social responsibility variable indicators, macroeconomic normal uncertainty, panel data processing, and micro-corporate governance and other aspects of non-financial indicators to establish a more reasonable and systematic financial crisis early warning model. In addition, from the empirical analysis of the model, we also found that Dantzig, another method of screening variables, has a good empirical effect. Although this article does not provide a systematic theoretical and algorithmic explanation, interested scholars can apply this method to other models or make predictions in the sample.

References

[1] Cristianini, N., & Shawetaylor, J. (2000). An Introduction to Support Vector Machines. China Machine Press.
[2] Fan, J., & Peng, H. (2004). On Nonconcave Penalized Likelihood With Diverging Number of Parameters. The Annals of Statistics, 32(3).
[3] Hearst, M. A., Dumais, S. T., Osman, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. IEEE Intelligent Systems & Their Applications, 13(4), 18-28.
[4] Kim, K. J. (2003). Financial time series forecasting using support vector machines. Neurocomputing, 55(1/2), 307-319.
[5] Lawrence, S., Giles, C. L., & Tsoi, A. C. (1997). Lessons in Neural Network Training: Overfitting May be Harder than Expected. Paper presented at the Fourteenth National Conference on Artificial Intelligence & Ninth Innovative Applications of Artificial Intelligence Conference.
[6] Min, J. H., & Lee, Y. C. (2005). Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. Expert Systems with Applications, 28(4), 603-614.
[7] Vapnik, V., & Vapnik, V. (1995). The Natural of Statistical Learning Theory.
[8] Vapnik, V. N. (1998). Statistical Learning Theory (Adaptive and Learning Systems for Signal Processing, Communications and Control Series).
[9] Cui J, Guo PJ, & Xia ZM. (2011). The properties of adaptive Lasso under Poisson log-linear regression model. Journal of Northwest University (Natural Science Edition), 41.
[10] Zhang, J. F. (2010). Research on SCAD based on simulation. Dalian University of Technology. Xiaomin Luo, Peng, D. T. (2020). Study on SCAD regular least square regression. Journal of Guizhou University (Natural Science Edition)(4).