

Hybrid PSO-SVM for Financial Early-Warning Model of Small and Medium-Sized Enterprises

Xinke Chong

College of Engineering and Applied Sciences, Stony Brook University, 11794, USA
Email: xinkechong92@outlook.com

ABSTRACT

As feature subset selection and parameter tuning are important for the performance of SVM-based models, a PSO-SVM model was provided which uses particle swarm optimization (PSO) to optimize both a feature subset and parameters of SVM simultaneously so as to improve the prediction result. Finally, the PSO-SVM model was applied to a financial early warning model, which shows a better performance than the pure SVM-based model.

Keywords: financial early warning model, particle swarm optimization (PSO), support vector machine (SVM)

1. INTRODUCTION

Financial early warning models are an important and extensively researched topic, because allow monitoring financial situations in real time, and they allow analysis and data mining, the acquisition of effective knowledge, and prediction of whether companies are in financial crisis[1-3]. They are beneficial to corporate management, investors, creditors and other corporate interests. Decision-making by stakeholders and management by the government and securities regulatory authorities are of important practical significance[4-6]. In order to obtain higher prediction accuracy, people apply statistical methods, artificial intelligence expert systems and theoretical methods to financial early warning models[7-9]. Considering that the final result of the company is a classification of whether the company is normal or in crisis, the company's financial early warning is similar to a binary decision-making model[10-14]. For example, Fitzpatrick and Beaver used a single financial ratio to show that two sample groups could be divided into bankruptcy and non-bankruptcy. On this basis, there are a lot of scholars who later used statistical methods to improve this classification, including linear discriminant analysis, multiple regression, logistic regression, etc. However, in traditional statistical methods, if the warning variables are required to be linear, normal, and independent, this limits the application of these methods in practice. An algorithm combining the improved discrete particle swarm (PSO) algorithm and the support vector machine (SVM) is proposed to select the optimal index set and use it for financial crisis early warning by Hu et al. This algorithm

is applied to the data of listed companies to test the early warning effect of the model 3 years in advance. Finally, it is compared with the model combining the common principal component analysis method and SVM. The result proves the rationality and superiority of the PSO-SVM model [1]. Ji et al. combined the actual current situation of our country with small and medium-sized enterprises, aiming at the shortcomings of the commonly used financial crisis early warning model, to effectively overcome the issues related to the small sample that existed in the traditional model of nonlinear and high-dimension problems such as local minimum point. Ji et al. proposed utilizing the classification ability of SVM (support vector machine) to establish a financial early warning model of the use of training and assessment of the financial data of listed companies, proving that the financial early warning model based on support vector machine (SVM) under the condition of small sample has good nonlinear modeling ability and generalization ability[4]. After joining the WTO, as the market gradually opened up to competition, our country's external environment related to enterprises is undergoing a series of profound changes which bring great opportunities, and also enlarge the risk of enterprise financial crises in our country. With a large number of state-owned enterprises and listed private companies, listed companies have become the main body of China's economy and the most representative part, therefore, in the presence of Chinese listed companies, strengthening the financial crisis early warning theory and empirical research is very important and urgent. First, research on financial crisis early-warning is

reviewed and summarized, mainly introducing the various common models and methods. Then, the methods of support vector machine and particle swarm optimization are introduced in detail. Wang puts forward a kind of improved particle swarm (PSO) algorithm in combination with the support vector machine (SVM) algorithm, to select the most optimal index set to be used for financial crisis early warning and the algorithm was applied to the latest data of listed companies. We check a model with a warning effect 3 years in advance. Finally, we use a principal component analysis (PCA) method combined with the SVM model and compare the results to the PSO-SVM model. The results prove the rationality and superiority of the PSO - SVM model [5]. Zhang et al used SPSS statistical software, using principal component analysis (PCA) to filter manufacturing companies' financial indicators, using MATLAB software with the help of particle swarm optimization (PSO), support vector machine (SVM) parameters optimization, and built a model based on PSO - LIBSVM corporate financial risk early warning model of the empirical analysis. This analysis showed that the model can be more accurate for manufacturing companies' financial risk measures. The artificial intelligence algorithm applied to effective attempts in the field of economic management, to analyze the company's financial risk has a certain realistic guiding significance[14].

With the development of new technologies, artificial intelligence expert system technology has been applied to early warning research. Among these methods, the support vector machine (SVM) method proposed by Vapnik using the principle of structural risk minimization has become the focus of attention [15, 16]. However, in the application of SVM in financial early warning, the choice of feature set and the selection of kernel function parameters have a great impact on the prediction results. This paper attempts to combine the particle swarm algorithm (PSO) and SVM to establish a hybrid model to determine the early warning indicators. The feature set and SVM parameters are optimized at the same time to improve the prediction accuracy of the SVM model.

2. INTRODUCTION TO PSO AND SVM

2.1. PSO

PSO is an optimization tool based on iteration. The system is initialized as a set of random particles (random solutions), and iteratively searches for the optimal value. In each iteration, the particles update themselves by tracking two "extreme values". One is the optimal solution found by the particle itself: this is called the individual extreme value P_{id} . The other extreme value is the optimal solution currently found by the entire population, and this extreme value is the global extreme value P_{gd} . When these two optimal values are found, the particle velocity and position are:

$$V_{id} = wV_{id} + c_1 rand() (p_{id} - x_{id}) + c_2 rand() (p_{id} - x_{id}) \quad (1)$$

$$X_{id} = X_{id} + V_{id} \quad (2)$$

Where: $rand()$ is a random number in the open interval (0,1); c_1 and c_2 are learning factors, usually $c_1 = c_2 = 2$ [17,18]; w is the inertia weight, which is used to control the influence of the velocity generated by the previous iteration on the velocity of this iteration. Generally, $w \in [0,1]$.

The velocity of the particle in each dimension will be limited to a maximum velocity V_{max} . If the updated velocity of a certain dimension exceeds the user-set V_{max} , then the velocity in this dimension will be limited to V_{max} .

PSO performs a certain random search according to the fitness value, and the entire search update process is the process of following the current optimal solution. Therefore, in most cases, all particles may converge to the optimal solution quickly.

2.2. SVM

SVM is a machine learning method based on statistical learning theory and the principle of structural risk minimization[19-21]. Its basic idea is to map samples in the input space to high-dimensional feature spaces through nonlinear transformations. In the feature space, one finds samples linearly separated by an optimal classification surface. The algorithm uses the classification interval to control the capacity of the linear learning machine, so as to minimize the structural risk and has a strong generalization ability under a limited sample.

SVM[22,23] is developed from the optimal classification surface in the case of linear separability. Set the equation of the classification surface as $x \cdot w + b = 0$, so that the linear separable sample set $(x_i, y_i) (i=1, 2, \dots, n), x \in R^d, y \in \{+1, -1\}$, the problem of finding the optimal hyperplane is equivalent to solving the quadratic programming problem under constraint conditions, and the optimization conditions are one of two types The distance between the two, and finally the optimal decision function is obtained:

$$f(x) = \text{sgn} \left[\sum_{i=1}^n a_i * y_i (x_i \cdot x) + b^* \right] \quad (3)$$

For nonlinear problems, input vectors need to be nonlinearly mapped to a higher-dimensional feature space, and then an optimal classification hyperplane can be constructed. This classification hyperplane can be defined as

$$f(x) = \text{sgn} \left[\sum_{i=1}^n a_i * y_i K(x_i \cdot x) + b^* \right] \quad (4)$$

Where: $K(x_i \cdot x)$ is the kernel function; the sign of $f(x)$ determines the category of x . Constructing the optimal hyperplane is equivalent to finding all non-zero points, and any data point corresponding to a non-zero point is the support vector of this optimal hyperplane .

3. PSO-SVM MODEL

Since the selection of feature sets and kernel function parameters have an important impact on the prediction performance of the SVM model, PSO is adopted in this paper to optimize the selection of feature sets and kernel function parameters at the same time to obtain close to the optimal feature set and kernel function parameters and improve the predictive ability of SVM models.

3.1. Feature set optimization

The selection of feature sets has filter and Wrapper modes [24]. The difference between them lies in whether the selection of feature sets is independent of the learning algorithm. If so, it is filter mode. otherwise, it is the wrapper method. In the way of the filter, feature selection is performed first, and then classification. Compared with the wrapper method, the filter method computation efficiency is higher. Wrapper training mode relies on input of a given feature set classification system and relies on testing the accuracy of the estimated classification. Although it is a slower process, the choice of feature set is usually the most optimal for classification.

In the study of financial early warning systems, the selection of feature sets is very important to improve the accuracy of the prediction. Moreover, when the number of features is large, the selection of feature sets is particularly important. This paper will adopt the wrapper method and use the PSO to optimize the algorithm selection to the optimal special collection of the SVM model.

The feature set is encoded in a binary string to represent some subset of the original feature set. Each bit in the string represents whether the corresponding feature is selected: 1 represents that the corresponding feature has been selected, and 0 represents the non-selection. Fig. 1 shows the encoding of the feature set.



Fig.1 Coding of feature set

Since the above codes cannot be optimized directly with PSO, the above binary codes can be converted into corresponding integers before optimization with PSO, and then the integers can be optimized with PSO.

3.2. Kernel parameter optimization

One of the biggest problems of SVM is how to choose the appropriate kernel function parameter values make the model have the best classification performance. However, it is not known in advance what value is the best parameter for a certain problem. Therefore, the optimization of kernel function parameters is a key factor to improve the prediction accuracy of the SVM-based financial early warning model.

In this paper, the RBF function is used as the kernel function of the SVM early warning model. Since the parameters involved play a crucial role in the performance of the SVM model, PSO is used to optimize these two parameters.

3.3. optimization of feature set and kernel function parameters with PSO

Fig.2 shows the workflow flow chart of the PSO-SVM model constructed in this paper.

The position of the particles is denoted by {V1, V2, V3}. V1 is the parameter C of the kernel function, V2 is the parameter δ^2 of the kernel function, and V3 is the feature set, and the fitness function is shown in Eq. (5). The feature set and kernel function parameter evaluation are carried out by an SVM process, and the process is repeated until the termination condition is met (this article uses the maximum number of iterations as the algorithm termination condition). The data set of the PSO-SVM model is divided into two parts: one part is used to train the SVM model; the other part is used to test the model. The fitness value of the particle is determined as follows:

$$Fitness = \sum_{i=1}^{Vsize} M_i N_{size} \tag{5}$$

Where: Vsize is the size of the test set, and M_i is the match between the predicted result and the actual result. When the predicted result is the same as the actual result, M_i is 1; otherwise, it is 0. It can be said that Fitness is the average classification of the SVM model Accuracy.

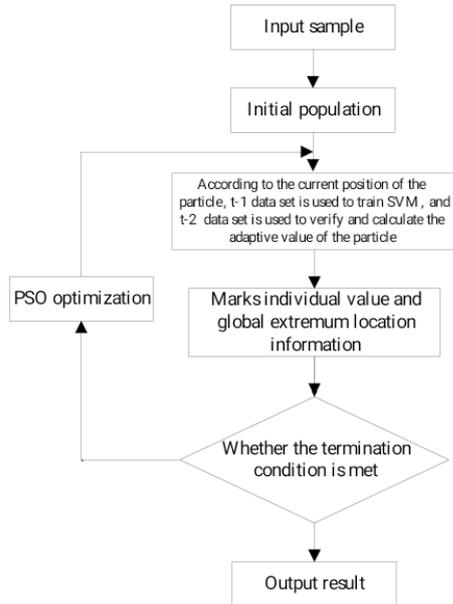


Fig. 2 The process of PSO-SVM model

4. EMPIRICAL RESEARCH

4.1. Sample selection

On March 16, 1998, the China Securities Regulatory Commission issued the "Notice on the Special Handling of Stocks During the Period of Abnormal Conditions of Listed Companies", requiring stock exchanges to implement special stock handling (ST) for listed companies with "abnormal conditions". Most of the ST stocks on the market are due to "two consecutive years of losses or net assets per share lower than the par value of the stock (1 yuan)", that is, the deterioration of financial indicators is the main reason for the special treatment of listed companies. Therefore, domestic research generally treats ST as a criterion for listed companies to fall into financial crisis. This article also regards ST companies as financial crisis companies and non-ST companies as normal companies.

Due to the limitations of obtaining information, the sample data in this article are from listed companies. The companies that performed normally from 2000 to 2005 and were first ST in 2006 are regarded as the failure group, and the company status is indicated by "1". This selection method can avoid packaging listed companies

into the failure group, which affects the accuracy of the model judgment. Since factors such as industry and company size have a greater impact on the financial indicators of the company, in order to control these interfering factors, this article selected according to the same industry and same size matching sampling of 2000 ~ 2006 years as the normal group, with the company's health expressed in "0".

By calculating the basic financial data of these sample companies at times t-1, t-2, t-3, t-4 (t is the year when the financial crisis began, this article refers to the year 2006 as t), 59 indicators are calculated. Long-term distribution observations have found that since t-2, there are multiple indicators showing a trend of increasing differences. Therefore, this article decided to use medium-term data as the solid sample collection of t-4, t-3, t-2, and t-1, in order to reflect the continuing change of financial data and cumulative effects. Companies with incomplete data were deleted. Finally, a total of 268 samples were selected (including 134 financial crisis companies and 134 matching companies). All basic data are From Wind Information (WIND).

4.2. Feature set selection

Combined with the research results at home and abroad, this paper initially selected 59 indicators in 7 aspects, including financial structure, solvency, growth ability, operating ability, profitability, cash flow, and others, in order to conduct a comprehensive analysis and systematically describe the company's situation.

The 59 indicators are selected through a variety of methods, and the results are shown in Table 1. In the table: "*" means selected; 59FS represents the feature set composed of all indicators; 7FS represents the selection from the feature set through the MDA stepwise method A feature subset composed of 7 indicators; 8FS represents a feature subset composed of 8 indicators selected from the feature set through the LR stepwise method; 38FS represents a feature subset composed of 38 indicators selected from the feature set through the independent sample t-test method ; "PSO-SVM" means that through the model proposed in this paper, a feature subset composed of 25 indicators is selected from the feature set.

Table 1 The feature subsets by different methods and the total predictors

Type	indicators	Index set obtained by different methods			
		7FS	8FS	38FS	39FS
Financial structure	Long-term debt to total capital ratio			*	*
	Long-term investment as a percentage of total assets			*	*

	Intangible assets and his share of assets		*	*	
	Proportion of receivables	*		*	
	Equity ratio		*	*	
	Long-term debt to working capital ratio			*	
Debt paying ability	Current ratio		*	*	*
	Liquidity ratio			*	*
	Number of times interest earned		*	*	*
	Tangible equity debt ratio			*	
	Asset-liability ratio		*	*	*
	Growth rate of return on equity		*	*	*
	Ebit per share growth rate		*	*	*
	Growth rate of net assets per share (adjusted)		*	*	
Growth ability	Growth rate of net assets per share		*	*	*
	Growth rate of earnings per share (net of earnings)	*	*	*	
	Growth rate of earnings per share		*	*	*
	Growth rate of undistributed earnings per share		*	*	
	Growth rate of capital accumulation fund per share			*	
	Main business profit growth rate			*	
	Main business revenue growth rate		*	*	
	Growth rate of total assets		*	*	
	Ratio of profits to cost		*	*	*
	Rate of stock turnover			*	*
	Turnover of fixed assets			*	
Business capacity	Turnover of account receivable			*	*
	Working capital/main business income			*	*
	Turnover of total capital		*	*	*
	Growth rate of shareholders' equity	*	*	*	*
	Return on equity		*	*	
	Net profit after deducting non-recurring gains losses		*	*	*
Profitability	Ebit per share	*	*	*	
	Net asset value per share		*	*	
	Net assets per share (adjusted)		*	*	
	Earnings per share		*	*	*
	Earnings per share (net of profits)		*	*	*

	Earnings per share			*	*	
	Undistributed profit per share			*	*	
	Capital reserve per share				*	
	Net profit margin on sales			*	*	*
	Gross profit margin			*	*	*
	Profit margin of main business			*	*	
	Net interest rate on assets			*	*	*
	Net cash from operations				*	
	Net cash to debt ratio			*	*	
	Growth rate of cash flow from operations per share				*	
Cash flow	Cash flow from operations per share				*	
	Growth rate of cash flow per share			*	*	
	Operating cash flow per share				*	
	Cash flow debt ratio				*	*
	Total cash debt to				*	
	Cash sales ratio			*	*	*
	Return on cash on assets			*	*	*
	Consumer price index	*	*		*	
	Company age				*	
	Company listing age	*	*	*	*	
	Company size			*	*	*
Rests	10 shareholders equity concentration degree	*	*	*	*	*
	10 liquid shareholders equity concentration degree			*	*	*
	SFC industry code				*	

4.3. Sensitivity analysis of feature sets and parameters in the SVM model

The data set is divided into a training set and a test set at a ratio of 1:1, and the SVM model is classified and predicted with different parameters and different feature sets. Fig. 3(a) shows that for different feature sets, the kernel function parameter $C=5$ represents the effect of $\delta 2$ on the prediction performance of the SVM model. Fig.

3(b) shows the effect of C on the prediction performance of the SVM model when the kernel function parameter $\delta 2 = 30$ for different feature sets. It can be seen from Fig. 3 that the SVM model's prediction performance is not only related to the selection of the feature set, but also related to the two parameters of the kernel function. Therefore, optimizing the parameters of the feature set and the kernel function at the same time will improve the prediction accuracy of the model.

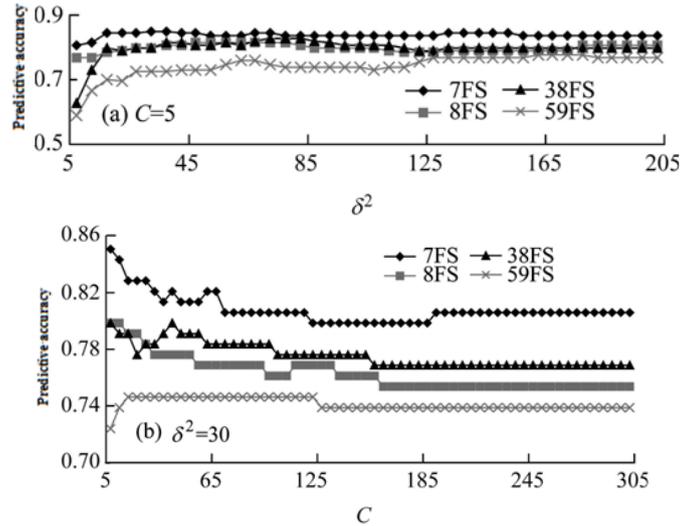


Fig.3 The accuracy of pure SVM models by different feature subsets

4.4. Analysis of the PSO-SVM model results

In this paper, the population size is 80 and the maximum number of iterations is 1000. The feature set and kernel function parameters are optimized simultaneously through the PSO-SVM model. The relationship between the model prediction accuracy and the number of iterations is shown in Fig. 4. The empirical results show that the PSO-SVM model optimized the kernel function parameters as $C=90.52$, and $\delta^2=12.93$.

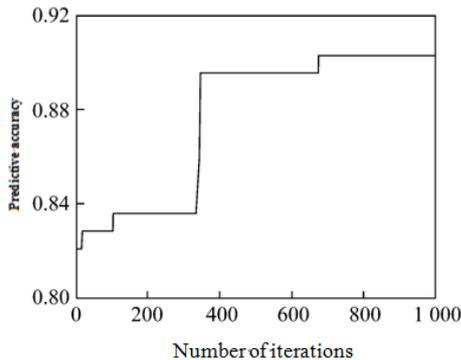


Fig.4 The relation between PSO-SVM's prediction accuracy and iterations

The results are shown in Table 1, and the optimal prediction accuracy of the model is 90.30%. The best prediction results are selected from the prediction accuracy obtained by the SVM model using different parameters and different feature sets, and combined with the prediction results of the PSO-SVM model, the prediction is accurate the prediction accuracy is shown in Table 2. It can be seen from Table 2 that the PSO-SVM model proposed in this paper automatically optimizes the selection of feature subsets and kernel functions through the PSO algorithm, while obtaining close to the optimal feature subsets and kernel functions. The PSO-SVM model also gets the best prediction results.

Table 2 The average prediction accuracy

Model	Feature subset	Accuracy / %
SVM	59	77.61
	38	82.84
	8	81.34
	7	85.08
PSO-SVM		90.30

5. CONCLUSION

This paper constructs an early warning model based on PSO-SVM and uses PSO to optimize the selection of the SVM model feature set and the kernel function parameters at the same time. The empirical results show that SVM, as a new machine learning method, has strong generalization ability. But the selection of feature set and kernel function parameters have a great impact on the prediction performance of the SVM model. The proposed PSO-SVM model optimizes the selection of feature sets and the SVM kernel function parameters at the same time, showing good results, improving the predictive power of the model.

REFERENCES

[1] HU D, Wang k, et al. Financial crisis warning model of listed companies based on PSO and SVM [J]. Journal of management, 2007, 4(5):588.
 [2] Wu L, Han H, Tang F. Multi-classification Financial early warning Model based on PSO-SVM [J]. Journal of Wuhan University of Technology (Information and Management Engineering Edition), 2013, 35(2):265-269.

- [3] Geng X. Research on Financial Crisis Early Warning of Listed Companies Based on Particle Swarm optimization algorithm and Support Vector Machine [D]. Nanchang University.
- [4] Ji C, Tang J. Discussion on SVM-based Financial Early-warning Model for Small and Medium-sized Enterprises [J]. *Market Modernization*, 2007, 000(11Z):327-328.
- [5] Wang K. Research on financial crisis early warning of listed companies based on PSO and SVM [D]. Anhui: University of Science and Technology of China, 2007. Doi:10.7666/d.y1077473.
- [6] Shen Q, Jiang J H , Jiao C X , et al. Modified particle swarm optimization algorithm for variable selection in MLR and PLS modeling: QSAR studies of antagonism of angiotensin II antagonists[J]. *European Journal of Pharmaceutical Sciences*, 2004, 22(2-3):145-152.
- [7] Cornec M. Concentration inequalities of the cross-validation estimator for Empirical Risk Minimizer[J]. *Mathematische Operationsforschung Und Statistik*, 2010, 51(1-3):43-60.
- [8] Neter J. Financial Ratios As Predictors Of Failure - Discussion[J]. *Journal of Accounting Research*, 1966, 4.
- [9] Altman E. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy[J]. *The Journal of Finance*, 1968, 23(4):589-609.
- [10] Altman E I, Haldeman R G, Narayanan P. ZETATM analysis A new model to identify bankruptcy risk of corporations[J]. *Journal of Banking & Finance*, 1977, 1(1):29-54.
- [11] Min J H, Lee Y C. Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters[J]. *Expert Systems with Applications*, 2005, 28(4):603-614.
- [12] Zavgren C. Assessing the vulnerability to failure of American industrial firms: A logistics analysis[J]. *Journal of Accounting Research*, 1985, 22:59-82.
- [13] Alifiah, Norfian M. Prediction of Financial Distress Companies in the Trading and Services Sector in Malaysia Using Macroeconomic Variables[J]. *Procedia Social & Behavioral Sciences*, 2014, 129(6):90-98.
- [13] Zhang J, Chen F, Wang B. Research on Financial Risk Early Warning of SVM Manufacturing Companies Based on PSO Optimization [J]. *Accounting Friends*, 2017, 000(014):52-56.
- [14] Fan A, Palaniswami M. Selecting bankruptcy predictors using a support vector machine approach[C] // *Proceedings of the International Joint Conference on Neural Network*. N J :Institute of Electrical and Electronics Engineers, 2000:354 -359.
- [15] Cao L. Support vector machines experts for time series forecasting[J]. *Neurocomputing*, 2003, 51:321-339.
- [16] Hu X. Introduction to particle Swarm optimization algorithm [EB/ O L].
- [17] Gao H, Zhou C, Gao L. General particle swarm optimization model[J]. *Chinese Journal of Computers*, 2005, 28(12):1980-1987.
- [18] Vapnik V N. The nature of statistical learning theory.[M]. N Y :Springer-Verlag, 1995.
- [19] Zhang X. Introduction to statistical learning theory and support vector machines[J] . *Acta Automatica Sinica*, 2000, 26(1):32-42.
- [20] Vapnik V N. An overview of statistical learning theory[J]. *Neural Network*, 1999, 10(5) :988 -999.
- [21] Deng N, Tian Y. A new method for data Mining: Support vector machine [M]. Beijing: Science Press, 2004.
- [22] Tay F, Cao L. Application of support vector machines in financial time series forecasting[J]. *Omega*, 2001, 29:309-317.
- [23] Sun Z, Bebis G, Miller R. Object detection using feature subset selection[J]. *Pattern Recognition*, 2004, 27 :2165-2176.