A Default Discrimination Method for Manufacturing Companies by Improved PSO-based LS-SVM

Weiwei Wang[1], Jie Cao[1] , Hongke Lu[2] and Jian Wang[3]

[1]*China institute of manufacturing development, Nanjing University of Information Science & Technology, Nanjing, China, 210044*
[2]*School of Economics & Management, Southeast University, Nanjing China, 210096*
[3]*Jiangsu Jinnong Information Co., Ltd., Nanjing China, 210019*

Cj@nuist.edu.cn, luhkseu@gmail.com

Abstract

Loan default evaluation and discrimination is a complicated issue because of its nonlinearity and uncertainty. Least square support vector machine (LS-SVM) has been successfully employed to solve regression and time series problem. This paper proposes a novel PSO-LS-SVM model based on the improved PSO algorithm to optimize parameters of LS-SVM, which is a new improved form by synthesizing the exiting model of PSO. Some evaluation indices, which are reduced without information loss by a genetic algorithm, are used to train PSO-LS-SVM and discriminate between healthy and default testing samples. A case study based on financial data acquired from listed companies has been carried out. Result has shown that the proposed model has a distinct improvement in the aspect of accuracy rate as compared to PSO-SVM, LS-SVM, SVM and BP neural network.

Keywords: *Least squares support vector machines (LS-SVM); Particle swarm optimization (PSO); Rough set; Loan default discrimination*

1. Introduction

Since late 1980s, with the deepening of financial liberalization and globalization, the volatility of financial markets gets intensified obviously. All banks and investors are facing more and more severe financial risks. Especially the sub prime mortgage crisis swept across the world in 2007, and it demonstrated that the financial operation pattern under globalization had changed fundamentally. All financial supervisory systems are facing unprecedented challenges.

The research by World Bank about the global bank crisis shows that credit risk is the main factor which leads to bank bankruptcy (Gu, Tang & Feng, 2006). In order to strengthen supervision of credit risk, the Basel council published two capital agreements. Measure and assessment of the enterprise credit default probability is the key element of the Internal Ratings-Based Approach (IRB) of new Basel Accord. Nowadays, more and more scholars have paid much attention to it. Exploring a more scientific and effective loan default discrimination method is a major topic of credit risk research (Ke & Feng, 2008). Meanwhile, international financial community and academic study mainly regard loan default discrimination problems as a classification problem of pattern recognition. According to financial and non-financial factors of the business community, researchers have summed up classification rules and established a credit default discrimination model.

Numerous methods have been put forward to construct a satisfactory loan default discrimination model to study the credit risk problem. Approaches for parameter selection can be categorized into two models. The first one is the statistics model including logistic method, mathematical programming, k nearest neighbor, Bayesian model and so on (Gerda, Marc & Geert, 2003; Arnold & Tomohiro, 2010).The other one is the artificial intelligence model such as neural network, genetic algorithm, decision tree and expert system (Du, 2010; Leardi, 2009; Huimin, 2007; Shu, 2005). Because of the complexity of credit risk and the abnormal structure of data, traditional pattern classification algorithms cannot meet the requirement of loan default discrimination any longer. Support vector machine (SVM), one of the new techniques for pattern classification, has excellent study performance and can solve the issues of small sample, nonlinear, high dimension and local minimum point (Senf, Chen, & Zhang, 2006). So it has been successfully and widely used in pattern recognition, regression estimation, probabilistic density estimation, time series forecasting and credit risk evaluation (Yu, Wang, & Cao, 2009; Li, Chen, & Wei, *et al.*, 2007; Tinghua, Houkuan, & Shengfeng, *et al.*, 2010; Shen & Zhen, 2009). Least square support vector machine (LS-SVM), the extension of SVM, has more advantages of SVM (Mathias & Mohamed, 2009; Wang & Guo, 2008). Extensive empirical study (Gestel, Suykens, Baesens, *et al.*, 2004) has shown that LS-SVM is comparable to SVM in terms of generalization performance. At the same time, it can overcome slow speed of large scale data for SVM and has the advantage of easy calculation and speedy (Suykens, Gestel, & De, 2002). Practice has shown that applying SVM model to credit evaluation of enterprises and individuals is well targeted and adapted (Xiao & Fei, 2006).

Nowadays, it is very difficult to select optimal parameters for LS-SVM model. For traditional approach, it is difficult to calculate and the result is not quite good. As we know, proper parameters setting can improve LS-SVM classification accuracy. In this paper, particle swarm optimization (PSO) is employed to select the penalty function C and the kernel function σ. Compared with other optimization methods including genetic algorithm, PSO is a new computation intelligence technique, which has been proved to be an optimization method with good performance. It is simple, fast and easy to be manipulated and realized with less parameters (Guo, Yang, Wu, *et al.*, 2008).The parameters of PSO are simple and they do not need complex adjustment. Thus, PSO is a useful approach to acquire optimal parameters of LS-SVM, which overcomes time consuming and blindness of cross validation. Additionally, it makes use of small sample training of LS-SVM, which is quite beneficial to data processing of loan default discrimination. The simulation result has demonstrated that PSO-LS-SVM model has better generalization ability and higher classification accuracy than that of PSO-SVM, LS-SVM, SVM and BP neural network.

2. Methodology

2.1 Least Squares Support Vector Machines

LS-SVM is a variant of SVM, which leads to solving linear KKT systems (Zhang & Zhang, 1999). LS-SVM differs from SVM where the quadratic programming is transformed to linear programming, which offers a simpler formulation and fewer computational requirements. In addition, the threshold b is calculated as a whole in the procedure（Suykens, Van Gestel, De Brabanter, *et al.*, 2002）.

Given a training set of instance-label pairs $\{x_i, y_i\}, i = 1, \cdots, n$, where $x_i \in R^l$, $y_i \in \{+1, -1\}$. The LS-SVM model in feature space is as follows:

$$y(x) = w^T \phi(x) + b \tag{1}$$

where the nonlinear mapping $\phi(\cdot)$ will map the input data to the high-dimensional feature space. The optimization problem of LS-SVM applies the approximate function:

$$\min_{w,b,\xi} J(w,\xi) = 1/2w^T w + (C/2)\sum_{i=1}^{n} \xi_i^2$$
$$subject\ to\ \ y_i = \phi(x_i) \times w + b + \xi_i;\ \ i = 1, \cdots, n \tag{2}$$

where $\xi_i \geq 0$ is the non-negative slack variable and $C > 0$ is a penalty parameter on the training error. The optimization problem will be solved by introducing Lagrange multiplier α_i.

$$L = \|w\|^2 + (C/2)\sum_{i=1}^{n} \xi_i^2 - \sum_{i=1}^{n} \alpha_i (\phi(x_i) \times w + b + \xi_i - y_i) \tag{3}$$

where α_i are Lagrange multipliers. According to the condition of KTT:

$$\partial L / \partial w = 0, \partial L / \partial b = 0, \partial L / \partial \xi = 0, \partial L / \partial \alpha = 0 \tag{4}$$

Then the following result can be obtained:

$$w = \sum_{i=1}^{n} \alpha_i \phi(x_i),\ \ \sum_{i=1}^{n} \alpha_i = 0,\ \ \alpha_i = C\xi_i,\ \ \phi(x_i)w + b + \xi_i - y_i = 0 \tag{5}$$

According to Eq.(5), by eliminating the parameters w and ξ, the linear equations are gotten as follows:

$$\begin{bmatrix} 0 & e^T \\ e & NN^T + C^{-1}I \end{bmatrix}_{(n+1)\times(n+1)} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \tag{6}$$

Where *e* is the vector ($\mathrm{n} \times 1$) whose element is 1. I is an N-dimensional identity matrix,; $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_n)^T$, $y = (y_1, y_2, \cdots, y_n)^T$, $N = (\phi(x_1)^T, \phi(x_2)^T, \cdots, \phi(x_n)^T)^T$.

According to Mercer, the Kernel Function is defined as $K(x_i, x_j) = \phi(x_i) \times \phi(x_j)$.

The linear decision function is gotten:

$$f(x) = \sum_{i=1}^{n} \alpha_i K(x_i, x_j) + b \tag{7}$$

Kernel Function is used widely including linear kernel, polynomial kernel and RBF kernel. In machine learning theory, the RBF kernel function has been proved to have better generalization performance (Cambell, 2002). So RBF kernel is chosen and its expression is described as follows:

$$K(x_i, x_j) = \exp(-\|x_i - x_j\| / 2\sigma^2) \tag{8}$$

Where σ is the parameter which shows the width of RBF kernel.

2.2 Sample Attributes Reduction

Attribute reduction, one of the most important topics in knowledge discovery, refers to the ability to maintain the same conditions and to delete irrelevant or redundant knowledge without information loss. However, it is a NP-Hard problem to select an optimal subset from attribute set. Existing reduction algorithms, mainly from the core of rough set, employ a heuristic search method to exploit minimum reduction. In the rough set theory, $\beta = pos_C(D_j) - pos_{C-\{c_i\}}(D_j)$ is employed to examine the importance of condition attribute $c_i \in C$ corresponding to decision attribute D_j. If $\beta = 0$, condition attribute c_i is considered to be redundant and can be deleted.

This study implements the genetic algorithm to find out the minimum reduction (Tao, Xu, Wang, *et al.*, 2003). Therein, a set of evaluation indices are defined as $C = \{c_1, c_2, ..., c_n\}$, Ω denotes the evaluation space. Chromosome is an n-bit binary string, with each binary bit corresponding to an evaluation index. If the value of the binary bit is 1, the corresponding evaluation index is selected, otherwise, it is deleted. In this way, each chromosome corresponds to an attribute subset of Ω. The algorithm's fitness function f is defined as follows (Ke, Feng, 2008):

$$f(\theta) = 1 - \frac{L_\theta}{n} + \frac{C_\theta}{(m^2 - m)/2} \tag{9}$$

Where L_θ denotes the number of evaluation indices chosen from chromosome θ, C_θ denotes the number of rows covered by chromosome θ in decision table, n denotes the number of evaluation indices, and m denotes training sample.

2.3 The Improved Particle Swarm Optimization

The particle swarm optimization (PSO) is a new computation intelligence technique, which is proved to be an optimization method with good performance. The global optimizing model proposed by Kennedy (Kennedy & Eberhart, 1995) is described as follows:

$$\begin{aligned} v_{k+1} &= v_k + c_1 r_1 (pbest_k - x_k) + c_2 r_2 (gbest_k - x_k) \\ x_{k+1} &= x_k + v_{k+1} \end{aligned} \tag{10}$$

where v_k is the velocity of particle i, representing the distance to be travelled by this particle from its current position; k is the number of iterations; x_k represents the particle position; w is the inertial weight; c_1 and c_2 are positive constant parameters; r_1 and r_2 are random functions in the range [0,1]; $pbest_k$ (local best solution) is the best position of the k-th particle and $gbest_k$ (global best solution) is the best position among all particles in the swarm.

In general, the parameters w, c_1, c_2, r_1, r_2 are important factors that influence the convergence of PSO. In order to solve the problem, the convergence factor λ and inertia weight ω are applied to improve the basic particle swarm optimization model. Xia and Dong propose a new improved form by synthesizing the existing model of PSO.

To search for an optimal solution, each particle changes its velocity according to the cognition and social parts as follows（Xia, Dong, & Du, 2007）:

$$v_{k+1} = \lambda \cdot [\omega(t) v_k + c_1 r_1 (pbest_k - x_k) + c_2 r_2 (gbest_k - x_k)] \quad (11)$$

$$x_{k+1} = x_k + v_{k+1}\omega(t) \quad (12)$$

where $\lambda = \dfrac{2}{\left|2 - c - \sqrt{c^2 - 4c}\right|}$, $w(t) = 0.9 - \dfrac{t}{T_{\max}} \times 0.5$.

Where c_1 indicates the cognition learning factor; c_2 indicates the social learning factor, r_1 and r_2 are random numbers uniformly distributed in $U(0,1)$, and $T_{\max}$ is the maximum iteration. With the increasing of t, parameter ω will decrease from 0.9 to 0.4 linearly. Meanwhile, in order to balance the effect of random, according to research (Shi & Eberhart, 1998), $c_1 = c_2 = 2$.

2.4 Parameter Selection by the Improved PSO

This study applies an RBF kernel function in the LS-SVM to obtain an optimal solution. Two major RBF parameters in LS-SVM including *C* and σ must be set appropriately. Parameter *C* represents the cost of the penalty and the choice of value for *C* has influence upon classification outcome. If *C* is too big, the classification accuracy rate is very high in the training phase, but very low in the testing phase. If *C* is too small, the classification accuracy rate is unsatisfactory. Parameter σ has a much greater influence on classification outcomes than *C*, because its value affects the partitioning outcome in the feature space. An excessively big value for parameter σ results in over-fitting, while a disproportionately small value leads to under-fitting (Pardo & Sberveglieri, 2005; Lin, Ying, Chen, *et al.*, 2008).

As penalty function *C* and kernel function σ will affect the performance of LS-SVM, it merely means different impact on different data sets. The optimal parameter will greatly improve accuracy of LS-SVM. Therefore, parameter selection is a key issue for successful application of algorithm. In this paper, our improved PSO algorithm is used for parameters selection. To implement our proposed approach, this research uses the RBF kernel function for the LS-SVM classifier because RBF kernel function can analyze higher dimensional data. The formulation can be given as below：

$$K(x_i, x_j) = \exp(-\left\|x_i - x_j\right\| / 2\sigma^2) \quad (13)$$

The fitness function is one of key factors to measure whether the improved algorithm is good or bad. In the PSO-LS-SVM, LS-SVM parameters are represented by every particle. The fitness to which the particle corresponds is the property of algorithm with this group of parameters. The mean absolute percentage error (MAPE) is employed as the fitness function:

$$f = 1/n \sum_{1}^{n} \left|(y_i - \hat{y}_i) / 2y_i\right| \quad (14)$$

Where n means the number of test sample, y_i is actual value, $\hat{y}_i$ is predicted value, and f means the value of fitness.

The algorithm will end when prediction error reaches a given value or iteration number reaches the maximum number of iteration. The process of parameter selection by the improved PSO is as follows.

Step 1: Initialize the particle swarm algorithm, the initial population $X(t)$ is made up of particle $x_1, x_2, ..., x_n$ which is from primitive space R^n randomly. The velocity matrix $V(t)$ is made up of the particle's initial velocity $v_1, v_2, ..., v_n$.

***Step 2*:** Define the fitness of particle $f = 1/n\sum_{1}^{n}\left|(y_i - \hat{y}_i)/2y_i\right|$.

Step 3: Compare current fitness $f(x_i)$ with the best fitness $f(pbest_k)$ of every particle, if $f(x_i) < f(pbest_k)$, then $pbest_k = x_i$; Compare current fitness $f(pbest_k)$ of all particles in population, if $f(x_i) < f(pbest_k)$, then the optimal solution is $pbest_k = x_i$.

Step 4: Compute and update the velocity of each particle according to Eq. (11) and Eq. (12) and new population $X(t+1)$ will be generated. Speed adjustment rules are as follows:

$$v_i = \begin{cases} V_{\max}, & v_i > V_{\max} \\ -V_{\max}, & v_i < -V_{\max} \end{cases} \tag{15}$$

Step5: Check the termination condition. If the number of iteration reaches the pre-determined maximum number of iteration, return the current best individual as a result; Otherwise, T=T+ 1 and return to step 2

Step 6: The optimal parameters σ and C of LS-SVM will be obtained through these steps.

3. Empirical Analysis

3.1 Attribute Reduction and Sample Data

In this paper, some manufacturers listed on Shanghai and Shenzhen stock markets are selected as samples to evaluate default status of manufacturing companies. Data is collected from website http://www.hexun.com/ from 2005 to 2007. After deleting abnormal samples, 188 samples are finally acquired. Therein, 94 companies are 'non-ST' companies with a financial position and low default risk. These companies are considered as 'good' companies and y=1. The remaining 94 companies are 'ST or *ST' with a bad financial situation and high default risk. These companies are referred to as 'bad' companies and y=-1. Meanwhile, if a 'bad' company is misclassified as 'good' company, we call it as a type I error. On the contrary, if a 'good' company is misclassified as 'bad' company, it is recorded as a type II error. For the purpose of enhancing the generalization ability and discrimination accuracy of the new model, a random sampling method is employed to divide the data set into two parts. Totally 112 companies, composed of 56 'good' companies and 56 'bad' companies, are selected as a training sample set, while the remaining 76 companies are used as a test sample set. The test sample set comprising of 38 positive samples and 38 negative samples are taken to testify prediction accuracy of the proposed PSO-LS-SVM. The experiments are carried out 3 times.

The evaluation attributes are mainly from domestic commercial banks, especially domestic third-party credit evaluation agencies. The financial indices system mainly comprises of four parts including debt paying ability, financial efficiency, capital working ability and potential as shown in Table 1.

Table 1. Financial Indices of Loan Default Discrimination System

Factors	Evaluation indices
Debt paying ability	Current ratio（C_1）; Quick ratio（C_2）; Cash ratio（C_3）; Current assets versus total debt ratio （C_4）; Asset-liability ratio（C_5）; Cash flow debt ratio（C_6）; Interest rate safeguard multiplier（C_7）
Financial efficiency	Return on net assets（C_8）; Return on total assets（C_9）; Dominant business ratio（C_{10}）; Net profit margin（C_{11}）; Ratio of Profits to Cost（C_{12}）
Capital working ability	Accounts receivable turnover（C_{13}）; Turnover of current assets（C_{14}）; Inventory turnover ratio（C_{15}）; Turnover of total capital（C_{16}）
Potential	Main business ratio（C_{17}）; Net profit growth rate（C_{18}）; Total assets expansion ratio（C_{19}）; Net capital increasing ratio（C_{20}）

Since the rough set algorithm can only deal with discrete attributes, continuous data should be discretized first. Referring to the criteria of discrete interval classification (Jiang & Yuan, 2003), we can get the result of discretization, then credit evaluation attributes are reduced by genetic algorithm, and the crossover probability $P_c = 0.8$, mutation probability $P_m = 0.05$. Then 3 minimum reductions are achieved as below: $\{C_1, C_5, C_7, C_8, C_{12}, C_{14}, C_{18}, C_{20}\}$,

$\{C_2, C_5, C_7, C_8, C_{12}, C_{15}, C_{18}, C_{20}\}, \{C_2, C_5, C_7, C_8, C_{12}, C_{15}, C_{17}, C_{20}\}$.

According to the formula of the minimum cluster ratio (Ke & Feng, 2008), one minimum reduction including 8 evaluation indices is achieved: $\{C_2, C_5, C_7, C_8, C_{12}, C_{15}, C_{18}, C_{20}\}$.Quick ratio, asset-liability ratio, interest rate safeguard multiplier, return on net assets, ratio of profits to cost , inventory turnover ratio, net profit growth rate and net capital increasing ratio are selected as model variables. Compared with key evaluation indices which are used by current domestic and international scholars to study loan default discrimination problem, these 8 reduced evaluation indices can reflect the credit status and debt paying ability of the companies.

Additionally, normalization processing of data before the PSO-LS-SVM training is crucial with the goal of speeding up model convergence and reducing the impact of imbalance in data capacity to the network classifier. We use linear differential analysis which is defined as $x_{ij}' = \dfrac{x_{ij} - \min_i}{\max_i - \min_i} \in [0,1]$, where $\max_i$, $\min_i$ denote maximum and minimal value of all sample data in attribute C_i respectively, x_{ij} denotes the $i-th$ attribute in the $j-th$ sample, and x_{ij}' denotes the data after being normalized.

3.2 Search the Best with Iteration

Our implementation platform is carried out on the Matlab7.1, a mathematical development environment, by extending the Libsvm Version 2.82 which is originally designed by Chang and Lin (Chang & Lin, 2008). According to non-linear characteristic of loan default discrimination data, the RBF kernel function is used. In PSO-LS-SVM model, the number of population size is 20 and the maximum number of iterations is 100. We set acceleration parameters $c_1 = c_2 = 1.3$, convergence parameter $\lambda = 0.729$, the inertia weight $\omega_{\max} = 0.9$, $\omega_{\min} = 0.1$. In our experiment, we compare the performance of our classifier with four other popular methods: (I) PSO-SVM, (II) LS-SVM, (III) SVM, and (IV) BP neural network.

3.3 Empirical Results

(1) Classification comparison of the first experiment

Classification results of the first experiment for testing samples are shown in Table 2. Among the 76 samples, 64 samples are classified correctly. Therein, 5 normal enterprises are misjudged as the default enterprise by PSO-LS-SVM, and misjudgment ratio is 13.16%. 7 default enterprises are misjudged as normal enterprise by PSO-LS-SVM, and misjudgment ratio is 18.42%. The average misjudgment ratio is 15.79%. From classification results of other four models, the average misjudgment ratio of PSO-SVM is 18.42%. The average misjudgment ratio of LS-SVM is 19.74% and that of SVM is 19.74%. BP neural network has the highest average misjudgment rate, which is 22.37%. Obviously, the classification precision of PSO-LS-SVM is higher than that of other four models.

Table 2. Classification Results of the First Experiment

Model	Original value	Predict Value Default	Predict Value Normal	Sample size	Misjudgment rate
PSO-LS-SVM	Default	31	7	38	18.42%
	Normal	5	33	38	13.16%
PSO-SVM	Default	27	11	38	39.29%
	Normal	3	35	38	7.89%
LS-SVM	Default	28	10	38	26.32%
	Normal	5	33	38	13.16%
SVM	Default	27	11	38	28.95%
	Normal	4	34	38	10.53%
BP neural network	Default	27	11	38	28.95%
	Normal	6	32	38	15.79%

Table 3. Classification Accuracy of Five Models（1）

Model	The First experiment			The Second experiment		
	Type Ⅰ error	Type II error	Error rate	Type Ⅰ error	Type II error	Error rate
PSO-LS-SVM	18.42%	13.16%	15.79%	15.79%	10.53%	13.16%
PSO-SVM	39.29%	7.89%	18.42%	21.05%	7.89%	14.47%
LS-SVM	26.32%	13.16%	19.74%	23.68%	7.89%	15.79%
SVM	28.95%	10.53%	19.74%	26.32%	5.26%	15.79%
BP	28.95%	15.79%	22.37%	28.95%	10.53%	19.74%

Table 4. Classification Accuracy of Five Models（2）

Model	The First experiment			The Second experiment		
	Type Ⅰ error	Type II error	Error rate	Type Ⅰ error	Type II error	Error rate
PSO-LS-SVM	26.32%	7.89%	17.11%	20.18%	10.53%	15.35%
PSO-SVM	28.95%	7.89%	18.42%	29.76%	7.89%	17.10%
LS-SVM	28.95%	7.89%	18.42%	26.32%	9.65%	17.98%
SVM	42.11%	5.26%	23.68%	32.46%	7.02%	19.74%
BP	39.47%	7.89%	23.68%	32.46%	11.40%	21.82%

(2) Classification Results

The discrimination process of the first experiment is given specifically in Table 2. Results of other two experiments can be acquired by the same approach. Detailed discrimination results are demonstrated in Table 3 and Table 4. The prediction precision of discrimination model is measured by Type Ⅰ and Type II errors.

From the classification results, Type Ⅰ error rate of PSO-LS-SVM is much lower than that of other 4 models. However, from Table 3 and Table 4, Type II error of PSO-LS-SVM is higher than that of other 4 models. Average classification error rate of PSO-LS-SVM is much lower than that of other 4 models. In other words, average classification accuracy of PSO-LS-SVM is the highest of these five models.

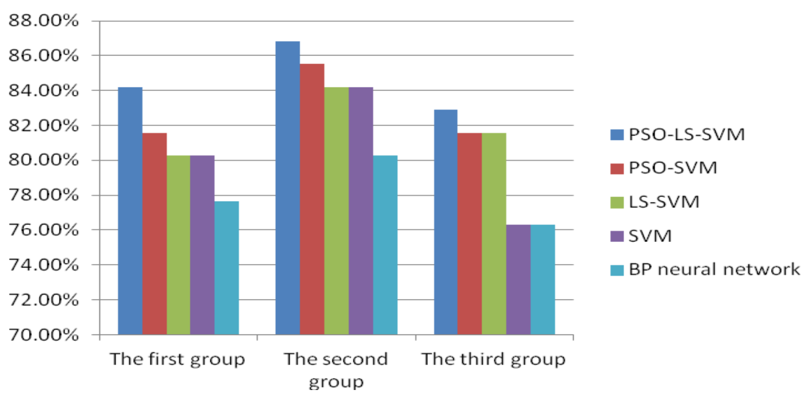

Figure 1. Classification Accuracy of Five Models in Three Groups of Experiments

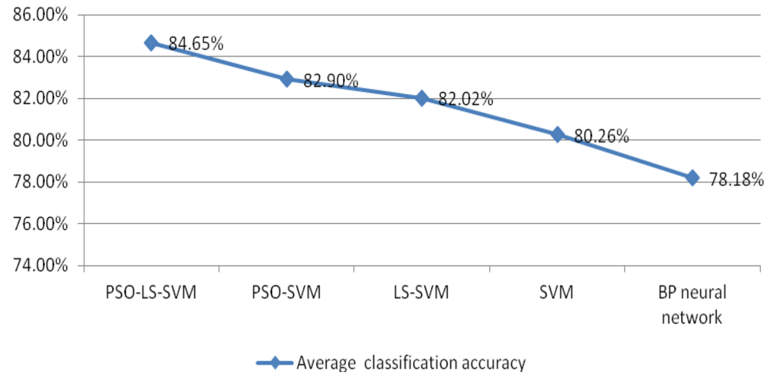

Figure 2. Average Classification Accuracy of Five Models in Three Groups of Experiments

In Figure 2, average classification accuracy of the proposed discriminate model is 84.65%. Average classification accuracy of PSO-SVM is 82.90%, and that of LS-SVM is 82.02%. For SVM, average classification accuracy is 80.26% and that of BP neural network is 78.18%. As shown in Figure 2, classification accuracy of PSO-LS-SVM is higher than that of other four discriminate models. Thus, the proposed model can obtain a better result in improving classification accuracy. And it definitely verifies that PSO-LS-SVM l is an effective tool in solving the problem of loan risk discrimination.

4. Conclusion

In this paper, an improved particle swarm optimization algorithm is put forward to select optimal parameters of LS-SVM. Over-fitting and under-fitting can be avoided effectively by optimizing LS-SVM parameters with improved PSO. Compared with genetic algorithm, algorithm parameters of particle swarm algorithm are simpler and do not involve complex adjustment. Thus, it is a useful approach to acquire optimal parameters of LS-SVM. The proposed model PSO-LS-SVM has been applied successfully to enterprises for loan default discrimination. Classification accuracy of PSO-LS-SVM model is higher than that of PSO-SVM, LS-SVM, SVM and BP neural network obviously. Empirical results indicate strong practicability of the proposed model and its high accuracy.

Acknowledgements

The authors would like to thank the anonymous referees for their valuable comments and suggestions. Their comments helped to improve the quality of the paper immensely. This work is partially supported by NSFC(60804047), the Science and Technology Project of Jiangsu province,china (BE2010201), Ministry of education, humanities and social sciences research project(11YJCZH005), Jiangsu provincial department of education philosophy and social science project(2010SJB790025) and the Priority Academic Program Development of Jiangsu Higher Education Institutions.

References

[1] Z. Arnold and A. Tomohiro, “A direct Monte Carlo approach for Bayesian analysis of the seemingly unrelated regression model”, Journal of Econometrics, vol. 159, no. 1, (**2010**), pp. 33-45.
[2] C. C. Chang and C. J. Lin, "LIBSVM: a library for support vector machines", http: //www.csie.ntu.edu.tw / ~ cjlin/libsvm/index.htm, (**2008**).
[3] K. L. Du, “Clustering: A neural network approach”, Neural Networks, vol. 23, no. 1, (**2010**), pp. 89-107.
[4] C. Gerda, A. Marc and M. Geert, “A quadratic bootstrap method and improved estimation in logistic regression”, Statistics & Probability Letters, vol. 61, no. 4, (**2003**), pp. 383-394.
[5] T. V. Gestel, J. A. K. Suykens, B. Baesens, S. Viaene, J. Vanthienen, G. Dedene, B. De Moor and J. Vandewalle, "Benchmarking least squares support vector machine classifiers", Machine Learning, vol. 54, no. 1, (**2004**), pp. 5-32.
[6] J. W. Gu, C. Y. Tang and Z. X. Feng, “Building default discriminating NN model on firm’s short-term loan-data-considering qualitative indexes and misclassification Loss”, Chinese Journal of Management Science, vol. 14, no. 15, (**2006**), pp. 104-108.
[7] X. C. Guo, J. H. Yang, C. G. Wu, *et al.*, "A novel LS-SVMs hyper-parameter selection based on particle swarm optimization", Neurocomputing, vol. 71, (**2008**), pp. 3211-3215.
[8] Z. Huimin, “A multi-objective genetic programming approach to developing Pareto optimal decision trees”, Decision Support Systems, vol. 43, no. 3, (**2007**), pp. 809-826.
[9] M. H. Jiang and X. C. Yuan, “Construction and application of GA-SVM model for personal credit scoring”, Journal of Hefei University of Technology, vol. 31, no. 2, (**2008**), pp. 267-283.
[10] J. Kennedy and R. C. Eberhart, “Particle swarm optimization”, Proc IEEE Int Conf Neural Networks. Piscataway: IEEE Press, (**1995**), pp. 1942-1948.
[11] K. L. Ke and Z. X. Feng, “Short-term loan default prediction based on integration of rough sets and genetic algorithm”, Systems Engineering-Theory & Practice, vol. 28, no. 4, (**2008**), pp. 27-34.
[12] K. L. Ke and Z. X. Feng, “Five-category classification of loan risk based on integration of rough sets and neural network system”, Control Theory &Application, vol. 25, no. 4, (**2008**), pp. 759-763.
[13] J. P. Li, Z. Y. Chen, L. W. Wei, W. X. Xu and G. Kou, "Feature selection via least squares support feature machine", International Journal of Information Technology and Decision Making, vol. 6, no. 4, (**2007**), pp. 671-686.
[14] S. W. Lin, K. C. Ying, S. C. Chen, *et al.*, “Particle swarm optimization for parameter determination and feature selection of support vector machines”, Expert Systems with Applications, vol. 35, no. 4, (**2008**), pp. 1817-1824.
[15] R. Leardi, “Genetic Algorithms”, Comprehensive Chemometrics, vol. 1, (**2009**), pp. 631-653.
[16] M. A. Mathias and C. Mohamed, “Model selection for the LS-SVM: Application to handwriting recognition”, Pattern Recognition, vol. 42, no. 12, (**2009**), pp. 3264-3270.
[17] M. Pardo and G. Sberveglieri, “Classification of electronic nose data with support vector machines”, Sensors and Actuators B Chemical, vol. 107, no. 2, (**2005**), pp. 730-737.
[18] Y. Shi and R. C. Eberhart, “Parameter selection in particle swarm optimization, Evolutionary Programming VII”, Proceedings of the Seventh Annual Conference on Evolutionary Programming, New York, (**1998**).
[19] J. A. K. Suykens, T. V. Gestel and B. J. De, "Least squares support vector machines", Singapore: World Scientific Pub. Co. Inc., (**2002**).
[20] J. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor and J. Vandewalle, "Least Squares Support Vector Machines", World Scientific, (**2002**).
[21] H. L. Shu, “Expert system methodologies and applications—a decade review from 1995 to 2004”, Expert Systems with Applications, vol. 28, no. 1, (**2005**), pp. 93-103.
[22] A. Senf, X. W. Chen and A. Zhang, "Comparison of one-class SVM and two-class SVM for fold recognition", Neural Information Processing, 13th Int Conf ICONIP 2006, Proc. Hong Kong, (**2006**), pp. 140-149.
[23] L. L. Shen and J. Zhen, "Gabor Wavelet Selection and SVM Classification for Object Recognition", Acta Automatica Sinica, vol. 35, no. 4, (**2009**), pp. 350-355.
[24] Z. Tao, B. D. Xu, D. W. Wang, *et al.*, "Rough Set Knowledge Reduction Approach Based on GA", Systems Engineering, vol. 21, no. 4, (**2003**), pp. 116-122.
[25] W. Tinghua, H. Houkuan, T. Shengfeng, *et al.*, “Feature selection for SVM via optimization of kernel polarization with Gaussian ARD kernels”, Expert Systems with Applications, vol. 37, no. 9, (**2010**), pp. 6663-6668.
[26] Y. Wang and W. Guo, “Local prediction of the chaotic fh-code based on LS-SVM”, Journal of Systems Engineering and Electronics, vol. 19, no. 1, (**2008**), pp. 65-70.
[27] W. B. Xiao and Q. Fei, “A Study of personal credit scoring models on support vector machine with optimal choice of kernel function parameters”, Systems Engineering-Theory & Practice, vol. 26, no. 10, (**2006**), pp. 73-79.

[28] K. W. Xia, Y. Dong and H. L. Du, “Oil Layer recognition model of LS-SVM based on improved PSO algorithm”, Control and Decision, vol. 22, no. 12, (**2007**), pp. 1385-1389.
[29] L. Yu, S. Y. Wang and J. Cao, “A modified least squares support vector machine classifier with application to credit risk analysis”, International Journal of Information Technology & Decision Making, vol. 8, no. 4, (**2009**), pp. 697-710.
[30] L. Zhang and B. Zhang, “Geometrical representation of McCulloch-Pitts neural model and its applications”, IEEE Transactions on Neural Networks, vol. 10, no. 4, (**1999**), pp. 925-929.