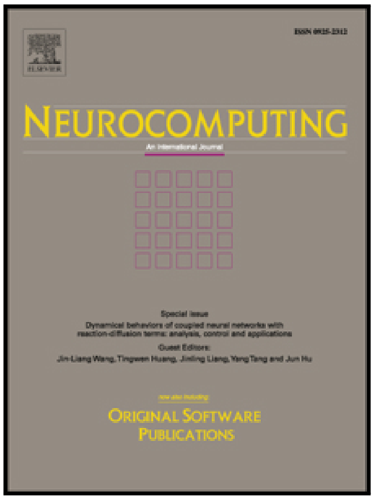
Communicated by Yukun Bao

Accepted Manuscript



An Improved SMO Algorithm for Financial Credit Risk Assessment–Evidence from China's banking

Qi Zhang, Jue Wang, Aiguo Lu, Shouyang Wang, Jian Ma

PII: S0925-2312(17)31232-8
DOI: 10.1016/j.neucom.2017.07.002
Reference: NEUCOM 18690

To appear in: *Neurocomputing*

Received date: 22 July 2016
Revised date: 26 June 2017
Accepted date: 4 July 2017

Please cite this article as: Qi Zhang, Jue Wang, Aiguo Lu, Shouyang Wang, Jian Ma, An Improved SMO Algorithm for Financial Credit Risk Assessment–Evidence from China's banking, *Neurocomputing* (2017), doi: 10.1016/j.neucom.2017.07.002

An Improved SMO Algorithm for Financial Credit Risk Assessment–Evidence from China's banking

Qi Zhang*‡, Jue Wang*†, Aiguo Lu§, Shouyang Wang*, Jian Ma‡
*Academy of Mathematics and Systems Science, Chinese Academy of Sciences, China
§Department of Applied Mathematics, Xi'an Shiyou University,Xian, 710065, China
‡Department of Information Systems, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong
†University of Chinese Academy of Sciences, Beijing 100190, China

Abstract—**With rapid development of financial services and products, credit risk assessment has recently gained considerable attention in the field of financial risk management. In this paper, an improved credit risk assessment approach is presented. Based on the credit data from China Banking Regulatory Commission (CBRC), a multi-dimensional and multi-level credit risk indicator system is constructed. In particular, we present an improved sequential minimal optimization (SMO) learning algorithm, named four-variable SMO (FV-SMO), for credit risk classification model. At each iteration, it jointly selects four variables into the working set and an theorem is proposed to guarantee the analytical solution of sub-problem. The assessment is made on China credit dataset and two benchmark credit datasets from UCI database and CD-ROM database. Experimental results demonstrate FV-SMO is competitive in saving the computational cost and outperforms other five state-of-the-art classification methods in credit risk assessment accuracy.**

Keywords: Credit risk assessment, SVM, Sequential minimal optimization (SMO), Four-variable working set

I. Introduction

The assessment of financial credit risk is emerging as an important research topic in the banking industry. The financial credit risk indicates the risk associated with financing, in other words, a borrower cannot pay the lenders, or goes into loan default. Credit risk assessment has become a particularly challenging issue for banks and financial institutions to access the performance of borrowers (customers), serving as the impetus to evaluate the credit admission or potential business failure of customers in order to make early actions. The great loss resulted from the financial distress or bankruptcy of customers usually leads to considerable criticism on the functionality of financial institutions due to the inappropriate evaluation of credit risk.

Most governments are forced to implement rescue plans for the banking systems with more effective credit risk assessment. In China, the massive credit boom poses challenge for the quality of bank assets. In fact, total bad loans reached 1.27 trillion yuan at the end of 2015, the highest since the global financial crisis, on the back of an economic slowdown and a ballooning corporate debt. An meticulous management information system is in urgent requirement. Credit risk assessment, which enables or supports an early-warning detection and fast response mechanism, is a key in this system. Since 2004, the China Banking Regulatory Commission (CBRC), which is responsible for regulation of banking industry in China, enables a reporting system for credit data collection. In recent years, CBRC has attached much importance to risk characteristics mining, custom's behavior analysis and risk assessment model.

Generally, credit risk refers to the risk that a bank borrower or a counterparty fails to meet its obligations in accordance with the agreed terms [1]. Numerous methods have been proposed in the literature to develop accurate classifier models to predict the default risk. Many statistic and optimization models are widely applied, such as linear discriminant analysis (LDA) [2], logistic regression analysis (LRA) [3], [4], multivariate adaptive regression splines (MARS) [5] and multi-criteria optimization classifier [6], [7]. However, the assumptions embedded within these statistical models, such as the multivariate normality assumptions for independent variables, are not satisfied in reality, which makes these methods theoretically invalid for finite samples [8]. Meanwhile, these models usually fail to capture enough information of nonlinear structure of real credit data. Recent studies focus on the research of artificial intelligent (AI) techniques for credit assessment, including artificial neural networks (ANN) [9], [10], radial basis function (RBF) model [11], decision tree [12], Bayesnet [13], extreme learning machine (ELM) [14], [15], support vector machine

(SVM) [16]-[19] and so on.

Specifically, SVM is a promising approach for credit risk evaluation [20]. It realizes the theory of VC dimension on principle of structural risk minimum and overcomes the over-fitting problem compared to artificial neural network. SMO algorithm developed by Platt [21] is one of the most efficient solutions for SVM training phase. It is derived by solving a series of small quadratic programming (QP) problems, where in each iteration only two variables are selected in the working set, as the small QP problems are solved analytically such as to avoid a time-consuming numerical QP method. The technique is popular and numerous efforts are made on improving and extending the classical model. For example, Song et al [22] put forward a new strategy by selecting several greatest violating samples set for the next several optimizing steps. Cai and Cherkassky [23] generalize Platt's SMO algorithm for SVM based multitask learning. Cao et al [24], [25] propose a parallel SMO which partitions the entire training data set into smaller subsets and then simultaneously runs multiple CPU processors to deal with each of the partitioned datasets. It's worth mentioning that Chen et al [26] study SMO-type decomposition methods using the two-element working set under a general and flexible way, which is called Chen-SMO in this paper and is a benchmark algorithm in this paper.

The above research make remarkale improvements, but they are all still limited to the two variables working selection proposed by Platt [21]. Thinking out of the framework, in the work of Lin et al [27], they generalize the traditional SMO algorithm to three-parameters SMO and the simulation results demonstrate their algorithm's superiority. According to these literatures, the training speed is a main limitation and important direction for making improvements of SVM algorithms. In this paper, we propose a novel and fast algorithm named four-variable sequential minimal optimization (FV-SMO). It is derived by solving a series of the QP problems with four variables at each iteration via the way of maximal violating pair (MVP). These QP problems are solved analytically so FV-SMO algorithm approaches the optimal solution more quickly to achieve the optimization goal. Moreover, a theorem is introduced on SVM-training to guarantee the existence of analytical solution of corresponding sub-problem. The proposed algorithm makes breakthrough in the training speed, algorithm complexity and generalization ability of SVM.

The proposed method is introduced to credit risk assessment. In this paper, we focus on large corporates with the loan more than 10 million RMB from the bank of China. At first, we construct a multi-dimensional and multi-level credit risk indicator system aiming to identify the most important credit risk indicators related to the hidden default risk by considering macroeconomic environment, enterprises' management ability and credit transaction behavior. China credit dataset is generated from CBRC monitoring system on the basis of credit risk indicator system. Two benchmark datasets, German and Darden credit datasets from UCI database and CD-ROM database respectively, are used to demonstrate the performance of the proposed method. In the numerical experiments, FV-SMO is compared with Chen-SMO in the computational cost and compared with five popular classification methods in credit risk assessment accuracy, including RBF, Multilayer-perception, Baysenet, decision tree, and Logisitic regression analysis. Experimental results show that FV-SMO is competitive in saving the computational cost and outperforms other credit assessment models.

This paper is organized as follows. Section 2 introduces SMO preliminaries. Section 3 presents the improved SMO algorithm based on four-variable working set. The credit risk indicator system and dataset generation process are shown in Section 4. Followed by the numerical experiments and result analysis in Section 5. Finally, the conclusion and future research makes up Section 6.

II. PRELIMINARY

A. Sequential Minimal Optimization

Consider the problem of separating the set of training vectors belonging to two classes: $D = \{(x_i, y_i)\}_{i=1}^{l}$, where l is the number of training samples, $x_i \in R^d$ is the ith training sample and $y_i \in \{+1, -1\}$ is the class label of x_i. SVM requires the solution of the following optimization problem:

$$\begin{aligned} \min\ & w(\alpha) = \tfrac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} y_i y_j \alpha_i \alpha_j K(x_i, x_j) - \sum_{i=1}^{l}\alpha_i \\ s.t\ & \sum_{i=1}^{l} y_i \alpha_i = 0 \\ & 0 \le \alpha_i \le C,\ i = 1, \cdots, l \end{aligned} \tag{1}$$

Sequential Minimal Optimization (SMO) is a simple algorithm that can quickly solve the SVM. It breaks the large QP problem into a series of smallest possible QP sub-problems, using the theorem from the work of Osuna et al [29] to ensure convergence. At every step, SMO chooses two Lagrange multipliers to jointly optimize, finds the optimal values for these multipliers, and updates the SVM to reflect the new optimal values. Suppose the two chosen variables are α_i,α_j, then the problem (1) can be written as the following optimization question:

$$\begin{aligned} \min\ & w(\alpha_i,\alpha_j) = \tfrac{1}{2}K_{11}\alpha_1^2 + \tfrac{1}{2}K_{22}\alpha_2^2 + y_1y_2K_12\alpha_1\alpha_2 \\ & -(\alpha_1+\alpha_2) + y_1\alpha_1\sum_{i=3}^{l} y_i\alpha_iK_{i1} + y_2\alpha_2\sum_{i=3}^{l} y_i\alpha_iK_{i2} \\ s.t\ \ & y_1\alpha_1 + y_2\alpha_2 = -\sum_{i=3}^{l} y_i\alpha_i = c \\ & 0 \le \alpha_i \le C,\ i = 1,\cdots,l \end{aligned} \tag{2}$$

if the origin solution is (α_1^0, α_2^0), then optimal solution can be presented as:

$$\begin{cases} \alpha_2^n = \alpha_2^0 + \frac{y_1y_2\nabla w(\alpha)_1 - \nabla w(\alpha)_2}{K_{11}+K_{22}-2K_{12}} \qquad (U \le \alpha_2^n \le V) \\ \alpha_1^n = \alpha_1^0 + y_1y_2(\alpha_2^0 - \alpha_2^n) \\ if \quad y_1 \ne y_2 : \\ U = max(0, \alpha_2^0 - \alpha_1^0), V = max(C, C + \alpha_2^0 - \alpha_1^0) \\ if \quad y1 = y2 : \\ U = max(0, \alpha_2^0 + \alpha_1^0 - C), V = max(C, \alpha_2^0 + \alpha_1^0) \end{cases} \tag{3}$$

The most important step of SMO is how to choose the working set. As pointed by Keerthi [31], Platt's [21] algorithm for the selection of working set can not guarantee the maximum degree of optimization of the objective function and they used a new method named Maximal Violating Pair to do the working set selection.

B. Working Set Selection

Currently, a very popular way to select the working set is "Maximal Violating Pair" (MVP) as follows:

$$\begin{cases} i \in \arg\max\limits_{t\in I_{up}(\alpha)} -y_t\nabla w(\alpha)_t \\ j \in \arg\min\limits_{t\in I_{low}(\alpha)} -y_t\nabla w(\alpha)_t \\ I_{up}(\alpha) = \{t|\alpha_t < C, y_t = 1\ or\ \alpha_t > 0, y_t = -1\} \\ I_{low}(\alpha) = \{t|\alpha_t < C, y_t = -1\ or\ \alpha_t > 0, y_t = 1\} \end{cases} \tag{4}$$

MVP can be derived through the Karush-Kuhn-Tucker (KKT) optimality condition of (1), to derive α^* for minimum $w(\alpha)$, it implies there exists a real number b^* and two nonnegative vectors λ^* and μ^* such that:

$$\begin{cases} \nabla w(\alpha^*) + b^*Y = \lambda^* - \mu^*, \\ \lambda_i^*\alpha_i^* = 0,\ \mu_i^*(C - \alpha_i^*) = 0, \\ 0 \le \alpha_i^* \le C,\ \lambda_i^* \ge 0,\ \mu_i^* \ge 0,\ i = 1,\cdots,l \end{cases} \tag{5}$$

where $\nabla w(\alpha) = Q\alpha - e$ is the gradient of $w(\alpha)$. The above condition can be rewritten as:

$$\begin{cases} \nabla w(\alpha^*)_i + b^*y_i \ge 0,\ if\ \alpha_i^* < C \\ \nabla w(\alpha^*)_i + b^*y_i \le 0,\ if\ \alpha_i^* > 0 \end{cases} \tag{6}$$

Since $y_i = \pm 1$, it can be derived from (6) that α^* is an optimal solution of (1) if and only $m(\alpha^*) \le M(\alpha^*)$.

III. An improved SMO algorithm based on four-variable working set

A. An important theorem

From above introduction of SMO, the key step of SMO is how to choose the two variables of the working set at each step. We propose an improved SMO algorithm named FV-SMO, the strategy is to choose four variables into the working set at each step. To advance the algorithm of FV-SMO, an important theorem is given in the following, this theorem guarantees the existence of optimal solution in our proposed algorithm FV-SMO.

Theorem 1. *Suppose* $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}$ *is a symmetry positive definite matrix, let* $x = (x_1, x_2)^T$ *and* $b = (b_1, b_2)^T$*, then the box constrained problem*

$$\min \quad q(x) = \frac{1}{2}x^TAx - b^Tx \tag{7}$$

$$s.t \quad l_i \le x_i \le u_i,\ l_i < u_i,\ i = 1, 2 \tag{8}$$

has a unique global optimal solution:

(I) $a_{12} \ge 0$,

$$\begin{cases} x_1^* = \min(\max(l_1, \overline{x_1}, \frac{b_1 - a_{12}u_2}{a_{11}}), \max(\frac{b_1 - a_{12}l_2}{a_{11}}, l_1), u_1) \\ x_2^* = \min(\max(l_2, \overline{x_2}, \frac{b_2 - a_{12}u_1}{a_{22}}), \max(\frac{b_2 - a_{12}l_1}{a_{22}}, l_2), u_2) \end{cases} \tag{9}$$

(II) $a_{12} < 0$,

$$\begin{cases} x_1^* = \min(\max(l_1, \overline{x_1}, \frac{b_1 - a_{12}l_2}{a_{11}}), \max(\frac{b_1 - a_{12}u_2}{a_{11}}, l_1), u_1) \\ x_2^* = \min(\max(l_2, \overline{x_2}, \frac{b_2 - a_{12}l_1}{a_{22}}), \max(\frac{b_2 - a_{12}u_1}{a_{22}}, l_2), u_2) \end{cases} \tag{10}$$

where $\overline{x_1} = \frac{b_1a_{22} - b_2a_{12}}{det(A)}$, $\overline{x_2} = \frac{-b_1a_{12} + b_2a_{11}}{det(A)}$.

The proof detail of Theorem 1 refers to Appendix A.

B. Solving the four-variable SVM subproblem

Assuming the working set of four-variables as $B = \{i_1, j_1, i_2, j_2\}$, and relatively the non-working set is $N = \{1,\cdots,l\} - B$, α, Q, e and Y can de decomposed:

$$\alpha = \begin{bmatrix} \alpha_B \\ \alpha_N \end{bmatrix} Q = \begin{bmatrix} Q_{BB} & Q_{BN} \\ Q_{NB} & Q_{NN} \end{bmatrix} e = \begin{bmatrix} e_B \\ e_N \end{bmatrix} Y = \begin{bmatrix} y_B \\ y_N \end{bmatrix} \tag{11}$$

The problem (1) is equivalent to the following sub-problem:

$$\begin{aligned}
&\min w(\alpha) = \tfrac{1}{2}\alpha_B^T Q_{BB}\alpha_B + (Q_{BN}\alpha_N - e_B)^T\alpha_B + const \\
&= \tfrac{1}{2}\left[\alpha_{i_1}\ \alpha_{j_1}\ \alpha_{i_2}\ \alpha_{j_2}\right]\begin{bmatrix} Q_{i_1i_1} & Q_{i_1j_1} & Q_{i_1i_2} & Q_{i_1j_2} \\ Q_{i_1j_1} & Q_{j_1j_1} & Q_{j_1i_2} & Q_{j_1j_2} \\ Q_{i_2i_1} & Q_{i_2j_1} & Q_{i_2i_2} & Q_{i_2j_2} \\ Q_{j_2i_1} & Q_{j_2j_1} & Q_{j_2i_2} & Q_{j_2j_2}\end{bmatrix}\begin{bmatrix}\alpha_{i_1}\\ \alpha_{j_1}\\ \alpha_{i_2}\\ \alpha_{j_2}\end{bmatrix} \\
&+(Q_{BN}\alpha_N - e_B)^T\begin{bmatrix}\alpha_{i_1}\\ \alpha_{j_1}\\ \alpha_{i_2}\\ \alpha_{j_2}\end{bmatrix} + const \\
&s.t\quad y_{i_1}\alpha_{i_1} + y_{j_1}\alpha_{j_1} + y_{i_2}\alpha_{i_2} + y_{j_2}\alpha_{j_2} = -y_N^T\alpha_N \\
&\qquad 0 \le \alpha_{i_1}, \alpha_{j_1}, \alpha_{i_2}, \alpha_{j_2} \le C
\end{aligned} \tag{12}$$

When solving (12), Chen et al in [26] study sequential minimal optimization type decomposition method under a general and flexible way of choosing the two-element working set, the iterative relationship is introduced as:

$$\alpha_i^{k+1} = \alpha_i^k - y_i d, \quad \alpha_j^{k+1} = \alpha_j^k + y_j d \tag{13}$$

Based on Chen's idea, we consider the iterative relationship:

$$\begin{aligned}
\alpha_{i_1}^{k+1} &= \alpha_{i_1}^k - y_{i_1}d_1, \quad \alpha_{j_1}^{k+1} = \alpha_{j_1}^k + y_{j_1}d_1 \\
\alpha_{i_2}^{k+1} &= \alpha_{i_2}^k - y_{i_2}d_2, \quad \alpha_{j_2}^{k+1} = \alpha_{j_2}^k + y_{j_2}d_2
\end{aligned} \tag{14}$$

so (12) is rewritten as:

$$\begin{aligned}
&\min w(\alpha) = \tfrac{1}{2}\left[d_1\ d_2\right]\begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22}\end{bmatrix}\begin{bmatrix} d_1 \\ d_2\end{bmatrix} - \left[b_1\ b_2\right]\begin{bmatrix} d_1 \\ d_2\end{bmatrix} \\
&s.t\quad l_1 \le d_1 \le u_1, \quad l_2 \le d_2 \le u_2
\end{aligned} \tag{15}$$

where

$$\begin{aligned}
a_{11} &= K_{i_1i_1} + K_{j_1j_1} - 2K_{i_1j_1} \\
a_{12} &= K_{i_1i_2} - K_{i_1j_2} - K_{j_1i_2} + K_{j_1j_2} \\
a_{22} &= K_{i_2i_2} + K_{j_2j_2} - 2K_{i_2j_2} \\
b_1 &= y_{i_1}\nabla w(\alpha^k)_{i_1} - y_{j_1}\nabla w(\alpha^k)_{j_1} \\
b_2 &= y_{i_2}\nabla w(\alpha^k)_{i_2} - y_{j_2}\nabla w(\alpha^k)_{j_2}
\end{aligned}$$

Note $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22}\end{bmatrix}$, Let A be a symmetric positive definite matrix. Given $y_{i_1} = y_{j_1} = 1$, since $0 \le \alpha_{i_1}^k - y_{i_1}d_1,\ \alpha_{j_1}^k + y_{j_1}d_1 \le C$, so $l_1 = \max(-\alpha_{j_1}^k, \alpha_{i_1}^k - C),\ u_1 = \min(C - \alpha_{j_1}^k, \alpha_{i_1}^k)$. For other values of $y_{i_1},\ y_{j_1},\ l_1,\ u_1$ has similar results. $l_2,\ u_2$ are available by the same analysis. By Theorem 1, problem (12) has the following optimal solution:

when $a_{12} \ge 0$,

$$\begin{cases} d_1^* = \min(\max(l_1, \overline{d_1}, \frac{b_1 - a_{12}u_2}{a_{11}}), \max(\frac{b_1 - a_{12}l_2}{a_{11}}, l_1), u_1) \\ d_2^* = \min(\max(l_2, \overline{d_2}, \frac{b_2 - a_{12}u_1}{a_{22}}), \max(\frac{b_2 - a_{12}l_1}{a_{22}}, l_2), u_2)\end{cases} \tag{16}$$

when $a_{12} < 0$,

$$\begin{cases} d_1^* = \min(\max(l_1, \overline{d_1}, \frac{b_1 - a_{12}l_2}{a_{11}}), \max(\frac{b_1 - a_{12}u_2}{a_{11}}, l_1), u_1) \\ d_2^* = \min(\max(l_2, \overline{d_2}, \frac{b_2 - a_{12}l_1}{a_{22}}), \max(\frac{b_2 - a_{12}u_1}{a_{22}}, l_2), u_2)\end{cases} \tag{17}$$

where $\overline{d_1} = \frac{b_1a_{22} - b_2a_{12}}{det(A)},\ \overline{d_2} = \frac{-b_1a_{12} + b_2a_{11}}{det(A)}$.

Finally, the solution of (12) is:

$$\begin{aligned}
\alpha_{i_1}^{k+1} &= \alpha_{i_1}^k - y_{i_1}d_1^*, \quad \alpha_{j_1}^{k+1} = \alpha_{j_1}^k + y_{j_1}d_1^* \\
\alpha_{i_2}^{k+1} &= \alpha_{i_2}^k - y_{i_2}d_2^*, \quad \alpha_{j_2}^{k+1} = \alpha_{j_2}^k + y_{j_2}d_2^*
\end{aligned} \tag{18}$$

C. An improved SMO algorithm based on four-variable working set

Here we employ the method of MVP to select the working set in the FV-SMO :

1)
$$\begin{aligned}
i_1 &\in \arg\max_{t \in I_{up}(\alpha)} -y_t\nabla w(\alpha)_t \\
j_1 &\in \arg\min_{t \in I_{low}(\alpha)} -y_t\nabla w(\alpha)_t \\
i_2 &\in \arg\max_{t \in (I_{up}(\alpha)\setminus\{i_1, j_1\})} -y_t\nabla w(\alpha)_t \\
j_2 &\in \arg\min_{t \in (I_{low}(\alpha)\setminus\{i_1,\ j_1,\ i_2\})} -y_t\nabla w(\alpha)_t
\end{aligned}$$

2) $B = \{i_1, j_1, i_2, j_2\}$

To sum up, the FV-SMO algorithm is motivated that multivariable coordinated optimization could reduce the number of iterations and training time. This method is derived by solving a series of the QP subproblems with four points and these subproblems are solved analytically. Thus, it can approach to the optimal solution much quickly, and further improve the performance of SMO-type learning algorithm greatly. We formally investigate the possible advantages by experiments analysis in the following section.The FV-SMO formal algorithm can be stated as follows:

Algorithm FV-SMO
Given dataset: $x_i, y_i,\ i = 1, 2, ...n$
Result: α_i solved by an analytical method
While it does not reach convergence, do:
step1: Given $\varepsilon > 0$ and $\alpha^0 = 0$. Set $k = 0$.
step2: If $m(\alpha^k) - M(\alpha^k) \leq \varepsilon$, stop; Otherwise by the above MVP to find a four-variable working set $B = \{i_1, i_2, j_1, j_2\}$.
Define $N \equiv \{1, \cdots, l\} - B$, α_B^k and α_N^k as sub-vectors of corresponding to B and N, respectively. And using formulas (16)-17) to get $d_1^*,\ d_2^*$, further by (14) derive optimal solution α_B^k.
step3: Gradient update: $\nabla w(\alpha^{k+1})$= $\nabla w(\alpha^k) + diag(Y)[(-K(:, i_1) + K(:, j_1))d_1^* + (-K(:, i_2) + K(:, j_2))d_2^*]$.
step4: Set $k = k + 1$ and go to step2.
end

IV. China's credit risk indicator system

As the economy skyrocketed in the past few years, China's financial system has grown exponentially. The assets managed by banks once grew more than 25% a year during the period of the massive fiscal stimulus plan to combat the global financial crisis. Only in the first two months of 2016, bank credit rose a significant 28% to RMB 3351 billion compared with the same period in previous year. Nowdays, China's economy is in the process of reform and structural adjustment, the banking institutions' risk management ability becomes increasingly important in the development.

Since 2004, CBRC has established a data collection system for monitoring the customers' loan behaviors monthly. Despite the data accumulated for more than ten years in the monitoring system, existing studies do not lend themselves to modeling the risk factors for big data and provide little guidance to policy makers in terms of loan application decision. Many experts point out that the key risk is coming from large corporate borrowing and from the reduction of profitability stemming from financial liberalization and heightened competition. In this paper, we focus on large corporates with loan more than 10 million RMB. We first construct a multi-dimensional and multi-level credit risk indicator system aiming to find the most important credit risk characteristics which will lead to the serious default risk, followed by the generation of China credit dataset.

A. *China credit risk indicator system*

First of all, the external factors are explored including macroeconomic, industry and region. The majority of China's credit is accumulated in fields of government infrastructure, real estate construction and large state-owned corporates. This leads to complicated causal relationship between credit risk, macroeconomic and monetary policy. It is reasonable to take macroeconomic factors as priority into consideration, such as GDP growth rate, M2 growth rate, interest rate and so on. Meanwhile, the overall situation of an industry closely relates to its credit behavior, three specific indicators are extracted: industry profit margin, concentration of loan investment and default rate of industry. According to different regional credit markets, implement differentiated regional credit policy also plays an important role in management process, so it is necessary to explore the regional dimensional factor. Regional industrial structure, regional economic development situation and regional default rate are investigated.

Second, we focus on the corporates' management ability. The enterprises' production, operation state and business behaviors directly affect the efficiency of credit funds use. The indicators related to operation situation, solvency and credit level are explored. Taking operation situation into consideration, corporates' capital scale and long-term viability of operation can be figured out. Generally, the operation situation can be expressed by measuring market capitalization, assets, cash flow and others. Here key financial indicators are investigated, such as asset scale, debt ratio, current ration and so on. Making the analysis of solvency promotes to clarify the ability of sustainable management and predict the future revenue, such as equity ratio, loan asset ratio are considered. For credit level, it is measured by the risk signals that given by the loan institutions as existing credit judgment towards the corporates, which could also contribute to the identity of default behavior in the future.

Third, the credit data from CBRC makes the corporates' transaction data mining possible. Two aspects including transaction behavior and association risk are mined. The customers' past loan behavior acts as the most convincing evidence for determining whether a customer should be granted good credit or not. The quantity and quality of loan, the lending bank information are explored as well. In addition, the credit interconnection among corporates are getting increasingly closer in recent years. The credit association promotes corporates to share capital, acquire excessive credit or escape from risk investigation. These behaviors increase the difficulty of credit

Figure 1: China credit risk indicator system

regulation and needs particular attention. Four major association relationship are investigated, including legal person, guarantee, stockholder and other business associates.

Finally, the whole indicator system is constructed with three overall dimension: external factors, management ability and trading behaviors. The second level includes 14 indicators including: macroeconomic, industry, region, operation situation, solvency, credit level, quantity, quality, bank, legal person, stockholder, business associates and guarantee. The third level is extended including 124 detailed indicators, such as regional default rate, loan bank concentration and so on. These indicators are not only extracted from the original data, but also derived from the analysis and mining of customer's transaction behavior. For instance, it is regarded as a risk signal if a customer often pay off the loan of one bank using the loan from other banks. Due to the limit of space, part of these indicators are listed in Figure 1.

B. *Generation of China credit dataset*

After the credit risk indicator system analysis, China credit dataset is extracted from the CBRC's credit data. Data preprocessing is implemented for converting the primary data to format. The techniques including cleaning, integration, transformation are used to process the dirty, incomplete and inconsistent data. Another important issue that needs to be clarified is the definition of default risk, we define a customer default if it is behindhand with its payment for more than three months. Once a default occurs in the credit history, the customer is marked as a positive sample.

The purpose of feature selection [32], [33], [34] is to filter out unrepresentative features from a given dataset, which is critical for a successful credit default classification model. For the credit risk indicator system as stated in Section 4, T-test and Wilcoxon signed ranks tests are used to distinguish indicators objectively. The criterion is whether the indicator changes significantly prior to the default occurs. 54 indicators are picked out by the single indicator test. Next, stepwise regression pares down the these indicators to eliminate the collinearity. Finally, 16 most representative indicators are chosen for the assessment model.

After data preprocessing and feature selection, the sample of China credit dataset has 60126 instances, 1822 default (positives) and 58324 non-default (negatives). The number of negatives is almost 32 times the size of positives. Since high imbalance of the data could seriously affects the model performance, downsample method is adopted to construct

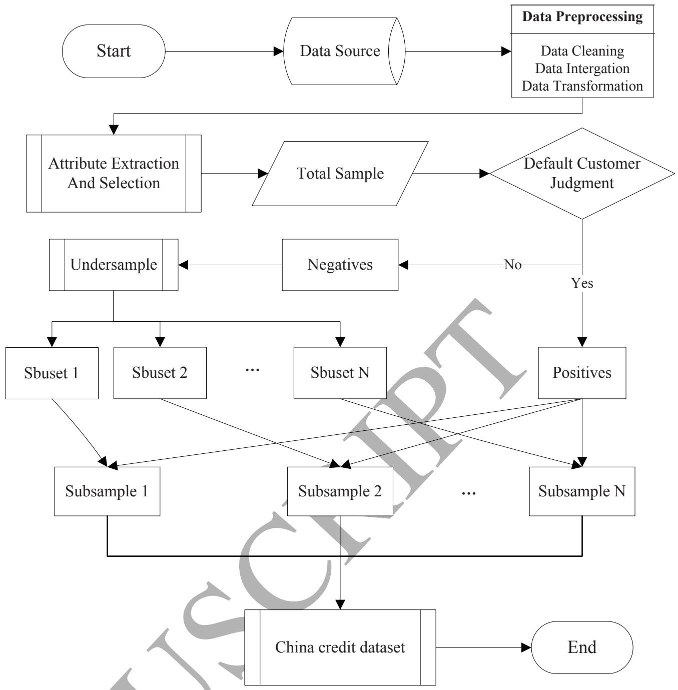


Figure 2: Flow chart for the generation of China credit dataset

subsample [35]. By one to one ratio, 1822 instances are picked out randomly from the sample of negatives. In order to account for the sampling error, we repeated sampling process for ten times and got ten subsamples for modeling. Figure 2 shows the generation process of the China credit dataset.

V. Numerical Experiments

A. *Dataset preparation*

Experiments are conducted on China credit dataset as stated in Section IV. China credit dataset contains 16 indicators and 3644 samples with positives and negatives balanced. Another two public datasets are also introduced for the credit risk assessment. German credit consists of 700 examples of creditworthy applicants and 300 examples where credit should not been extended. For each applicant, 24 indicators describe the individual credit history, age, loan amount, account balances, loan purpose, job title, and so on.

Table I: Description of testing datasets

Datasets	Number	Negative	Positive	Indicators
China credit	3644	1822	1822	16
German credit	1000	700	300	24
Darden credit	132	66	66	24

For Darden corporate credit dataset, it is from the CD-ROM database and includes 132 companies (66 non-risk cases and 66 risk cases). A total of 25 financial variables are computed for each of the 132 companies using data from the Compustat and from the Moodys Industrial Manual. The information of all datasets is shown in TableI.

B. Complexity comparison

To assess the complexity efficiency of the FV-SMO algorithm, RBF kernel- $K(x,y) = \exp(-\gamma * \|x - y\|^2)$ is chosen. All experiments are run on PC (Intel(R)core(TM)5/RAM16.0GB) in MATLAB.2016. The stopping condition ε, and the hyper parameters of C and γ need to be given. A superior algorithm should have stable and better performance with the changing parameters and different stopping condition. In Table II, C and γ are set at 1, the stopping condition ε ranges from 0.1 to 1×10^{-10}. In Table III, γ is set at 1 and ε is set at 1×10^{-10}.

C ranges from 0 to 5. From the results of Table II and Table III, the number of FV-SMO's iterations is obviously fewer than Chen-SMO. Student T-test and Wilcoxon's signed rank test are taken to identify the statistical significance of the results comparison. All the P values of experiments are statistical significant which demonstrate the superiority of FV-SMO.

As m-M can reflect the optimizing convergence rate for current iterating rate more intuitive, Figs.3-5 depicts the curves of m-M versus the iterating steps. The curves with blue and red colour are the results of FV-SMO and Chen-SMO, the C and γ are set at 1, ε is set at 1×10^{-10} . In most iterating steps, we can see the declines of objective m-M in FV-SMO are superior to that in Chen-SMO. The convergence speed of FV-SMO is significantly faster than that of Chen-SMO, which once again demonstrates the proposed FV-SMO outperforms Chen-SMO in the sense of faster convergence .

The results of classification accuracy are shown in Table IV-VI. Firstly, the Total accuracy of FV-SMO generally outperforms other classifiers for both China and German datasets, followed by MLP for China dataset and Bayesnet for German dataset. But for Darden dataset, FV-SMO is second only to Logistic. In terms of Type1 accuracy, FV-SMO is superior to other classifiers for China and Darden datasets, ranking second (0.453) on German dataset, the best is Bayenet (0.483). Then from the Type2 accuracy viewpoint, RBF has the best performance (0.887) for China dataset and FV-SMO has the best performance (0.903) for German dataset. FV-SMO has a relatively poor performance (0.622) compared to best result of Logisitic (0.819) for Darden dataset.

Figure 3: The change of m-M with iterating steps: China credit

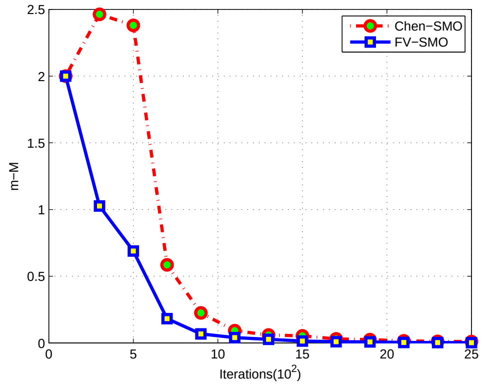


Figure 4: The change of m-M with iterating steps: German credit

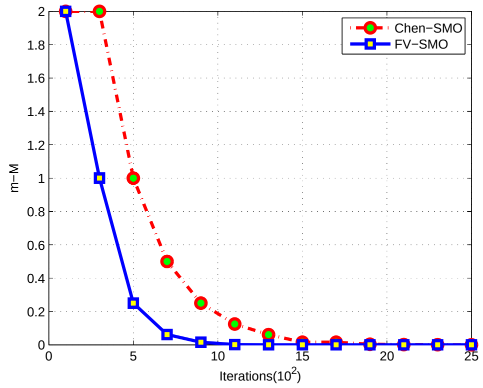


Figure 5: The change of m-M with iterating steps: Darden credit

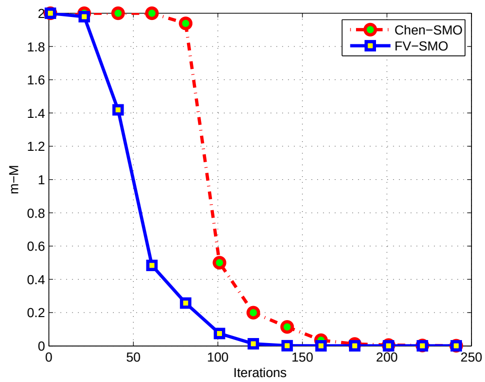


Table II: Executing iterations with changing ε on the three datasets

C=1, $\gamma = 1$	China credit		German credit		Darden credit	
ε	Chen-SMO	FV-SMO	Chen-SMO	FV-SMO	Chen-SMO	FV-SMO
0.1	233	105	1128	574	102	56
0.01	523	255	1669	845	143	83
0.001	983	519	2069	1058	195	103
1×10^{-4}	1342	691	2575	1258	249	121
1×10^{-5}	1917	860	2927	1478	295	139
1×10^{-6}	2217	1056	3230	1661	339	159
1×10^{-7}	2868	1374	3591	1817	383	178
1×10^{-8}	3346	1495	3979	1971	433	200
1×10^{-9}	4090	1780	4374	2172	469	222
1×10^{-10}	5012	2502	4785	2427	516	241
T-test (P value)	1.00×10^{-3}***		2.22×10^{-5} ***		1.20×10^{-4} ***	
Wilcoxon test (P value)	0.0890*		0.0058***		0.0110***	

*** represents 1% level significant, * represents 10% level significant.

Table III: Executing iterations with changing C on the three datasets

$\gamma = 1, \varepsilon = 1 \times 10^{-10}$	China credit			German credit			Darden credit	
C	Chen-SMO	FV-SMO	C	Chen-SMO	FV-SMO	C	Chen-SMO	FV-SMO
0.01	203	104	0.01	3776	1980	0.01	61	31
0.05	354	164	0.02	3831	2017	0.5	203	105
0.1	538	235	0.05	4018	2167	0.7	228	112
0.2	1086	432	0.1	4060	2010	0.9	416	196
0.3	2508	1049	0.2	4094	2119	1	516	214
0.5	3554	1514	0.5	4398	2166	1.5	758	348
0.8	4295	1886	0.8	4377	2298	2	817	368
1	5012	2502	1	4785	2427	3	854	393
1.5	5730	2448	1.5	6476	3322	3.5	873	405
2	6223	2501	2	6849	2501	5	944	387
T-test (P value)	3.20×10^{-2}***			4.91×10^{-7}***			4.78×10^{-4}***	
Wilcoxon test (P value)	0.0757*			1.83×10^{-4}***			2.57×10^{-2}***	

*** represents 1% level significant, * represents 10% level significant.

C. Accuracy comparison

Classification accuracy is the basic and decisive aspect in choosing the credit classification model. In order to check the performance of FV-SMO, we compare FV-SMO with five popular classification approaches in classification accuracy. The benchmark approaches include RBF, Multiplayer-perception, Bayesnet, Decision tree, and Logistic. For each approach, we adjusted the parameters to achieve the best classification accuracy. The five models are run using Weka 3.6. Given a classifier and an instance, there are four possible outcomes: if the instance is positive and it is classified as positive, it is counted as a true positive (TP); if the instance is negative and it is classified as positive, it is counted as a false positive (FP), the definition of TN and FN is the same. Six well known accuracy criteria are used to evaluate the performance of the classifier as follows:

(i) The total classification accuracy rate

$$Total\ accuracy = \frac{TP+TN}{TP+FP+TN+FN}$$

(ii) The identification rate of "bad" creditors

$$Type1\ accuracy = \frac{TP}{TP+FN}$$

(iii) The identification rate of "good" creditors

$$Type2\ accuracy = \frac{TN}{TN+FP}$$

(iv) How accurately of "bad" creditors have been classified

$$Precision = \frac{TP}{TP+FP}$$

(v) The mixed measure of classification

$$F1 - measure = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}}$$

In practical credit risk management, Type1 accuracy measures the identification rate of "bad" creditors, which means that the potential customer who is actually un-creditworthy is denied credit. The Type2 accuracy measures the identification rate of "good" creditors, which means a creditworthy customer is granted by the decision maker. Precision measures what the fraction is correctly categorised in all the positive predictions. F1-measure is a mixed criteria with a combination of Precision and Recall. Clearly, higher of these criteria corresponds to better predictive performance of classifier.

The receiver operating characteristic curve (ROC) [36] is a two-dimension graph in which true positive rate is plotted on the Y-axis and false positive rate is plotted on the X-axis. AUC based on the area under the ROC curve is another measure of classfier. Generally, a model with a larger AUC will have a better average performance.

Figure 6: ROC comparison for different models in China dataset

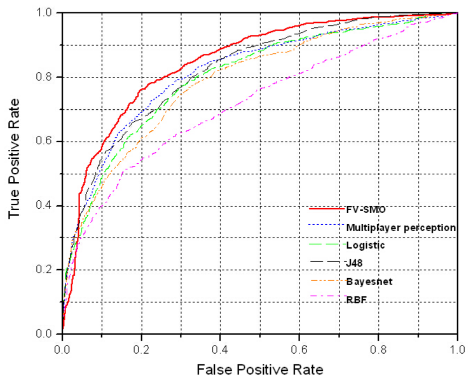


Figure 7: ROC comparison for different models in German dataset

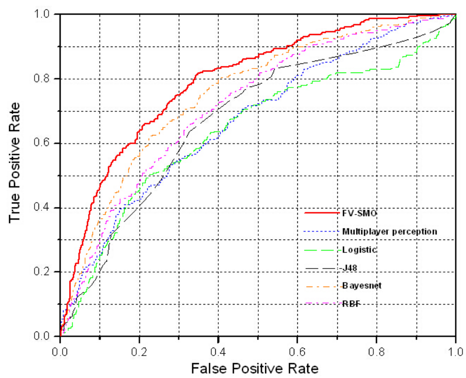


Figure 8: ROC comparison for different models in Darden dataset

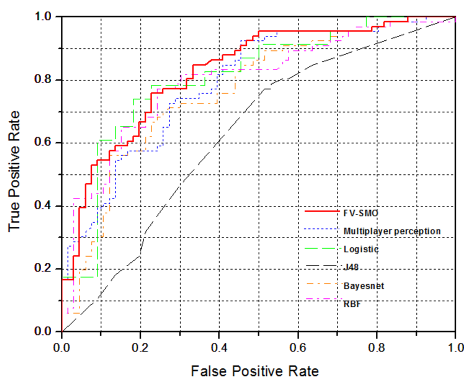


Table IV: Performance comparison for different models in China dataset

Model	Total Accuracy(%)	rank	Type1 Accuracy(%)	rank	Type2 Accuracy(%)	rank
RBF	0.653	6	0.419	6	**0.887**	1
Multilayer-perception	0.75	2	0.741	2	0.759	5
Bayesnet	0.705	5	0.608	5	0.803	2
J48	0.734	3	0.695	4	0.773	4
Logistic	0.729	4	0.709	3	0.749	6
FV-SMO	**0.778**	1	**0.768**	1	0.788	3
	Precision(%)	rank	F1-measure(%)	rank	ROC curve space	rank
RBF	**0.788**	1	0.547	6	0.718	6
Multilayer-perception	0.755	3	0.748	2	0.845	2
Bayesnet	0.755	4	0.673	5	0.78	5
J48	0.754	5	0.723	4	0.825	3
Logistic	0.739	6	0.724	3	0.803	4
FV-SMO	0.784	2	**0.776**	1	**0.849**	1

Table V: Performance comparison for different models in German dataset

Model	Total Accuracy(%)	rank	Type1 Accuracy(%)	rank	Type2 Accuracy(%)	rank
RBF	0.717	3	0.43	3	0.84	5
Multilayer-perception	0.709	4	0.29	5	0.889	3
Bayesnet	0.735	2	**0.483**	1	0.843	4
J48	0.692	6	0.217	6	0.896	2
Logistic	0.701	5	0.45	4	0.809	6
FV-SMO	**0.768**	1	0.453	2	**0.903**	1
	Precision(%)	rank	F1-measure(%)	rank	ROC curve space	rank
RBF	0.535	3	0.477	3	0.719	3
Multiplayer-perception	0.527	4	0.374	5	0.671	5
Bayesnet	0.569	2	0.523	2	0.755	2
J48	0.471	6	0.297	6	0.676	4
Logistic	0.502	5	0.475	4	0.641	6
FV-SMO	**0.667**	1	**0.54**	1	**0.794**	1

Table VI: Performance comparison for different models in Darden dataset

Model	Total Accuracy(%)	rank	Type1 Accuracy(%)	rank	Type2 Accuracy(%)	rank
RBF	0.742	3	0.848	2	0.742	2
Multiplayer-perception	0.689	5	0.758	4	0.621	5
Bayesnet	0.72	4	0.712	6	0.727	3
J48	0.629	6	0.788	3	0.47	6
Logistic	**0.778**	1	0.739	5	**0.819**	1
FV-SMO	0.743	2	**0.864**	1	0.622	4
	Precision(%)	rank	F1-measure(%)	rank	ROC curve space	rank
RBF	0.767	3	**0.806**	1	0.821	2
Multiplayer-perception	0.667	5	0.709	5	0.791	4
Bayesnet	0.723	4	0.718	4	0.766	5
J48	0.598	6	0.68	6	0.63	6
Logistic	0.81	2	0.77	3	0.808	3
FV-SMO	**0.864**	1	0.773	2	**0.826**	1

For the measurement of Precision and F1-measure, FV-SMO also yeilds a very good performance. Precision of FV-SMO ranks the second with a tiny gap (0.004) behind RBF on China dataset. F1-measure of FV-SMO ranks the second behind Logistic on Darden dataset. For other comparisons, FV-SMO ranks the first. The area under the receiver operating characteristic ROC curve is applied as another performance measurement. Figure 6 - 8 show the performance of the ROC curve for the three datasets. It is obvious that FV-SMO outperforms other models in terms of ROC.

Comparing with the empirical results of the three datasets, it is seen that Type2 accuracy is better than Type1 accuracy for China and German datasets, which means it is more difficult to catch the "bad" creditors from all the applicants, especially for the unbalanced dataset of German. But the result is inconsistent on Darden dataset. One possible reason is that different credit markets have different credit characteristics and the other is more nonlinearity in China and German datasets than that in Darden dataset.

Meanwhile, there is an interesting finding from Type1 accuracy and Type2 accuracy. For example, RBF ranks the first of Type2 accuracy (0.887), but performs the worst of Type1 accuracy (0.419) on China dataset. For German dataset, MLP and J48 have a poor performance in terms of Type1 accuracy, only 0.29 and 0.21, but get quite high performance of Type2 accuracy, 0.889 and 0.896, ranking top three with a slight difference to the FV-SMO. It is shown that some classifiers have the tendency to get a high recognition rate of majority class by predicting most samples as the "good" ones, especially on the imbalanced dataset, making the classifiers unsuitable for the credit risk assessment. From the above analysis, it can be concluded the proposed FV-SMO is promising in comparison with the other five popular classification approaches.

VI. Conclusion and future work

In this paper, we present a novel SMO learning algorithm on a four-variable working set for classification model and applied it to China credit dataset and two benchmark datasets. This method derived by solving a series of the QP sub-problems with four variables and these sub-problems are solved analytically so that the proposed method approaches to the optimal solution more quickly. Numerical results demonstrate that the proposed method has faster speed with statistical significance. Besides, experimental results also illustrate that FV-SMO can get the satisfactory performance in the classification accuracy, which provides compelling evidence of the advantages of FV-SMO. Given its encouraging performance, we are aiming to extend the algorithm to solve the problem of multi-class and regression problem instead of the binary classification.

Another contribution of this work is the multi-dimensional and multi-level credit risk indicator system. According to our knowledge, it is the first attempt to build the comprehensive indicator system on real credit data of China's banking. The system can not only help the banking managers and the audience of this paper to understand the overall situation of China's credit risk, but also scree out the key indicators that should been monitored by the policy makers. For future work, we could explore this system for more credit risk management applications.

VII. Acknowledgements

This work is supported by Chinese Academy of Sciences (CAS) Foundation for Planning and Strategy Research(KACX1-YW-0906), Youth Innovation Promotion Association of CAS, and the National Natural Science Foundation of China (NSFC No.71271202).

References

[1] Basel Committee on Banking Supervision . Principles for the Management of Credit Risk - final document. 2000.

[2] G. Karels, A. Prakash. Multivariate normality and forecasting of business bankruptcy. Journal of Business Finace Accounting. 14 (4) (1987) 573-593.

[3] L.C. Thomas. A survey of credit and behavioral scoring: Forecasting financial risks of lending to customers. International Journal of Forecasting. 16 (2) (2000) 149-172.

[4] D. West. Neural network credit scoring models. Computers and Operations Research. 27 (2000), 1131-1152.

[5] J.H. Friedman. Multivariate adaptive regression splines. The Annals of Statistics. 19 (1) (1991) 1-67.

[6] X. Zhu, J. Li, D. Wu, H. Wang, C. Liang. Balancing accuracy, complexity and interpretability in consumer credit decision making: A C-TOPSIS classification approach. Knowledge-Based Systems. 52 (2013) 258-267.

[7] Z. Zhang, G. Gao, Y. Shi. Credit risk evaluation using multi-criteria optimization classifier with kernel, fuzzification and penalty factors. European Journal of Operational Research. 237 (1) (2014) 335-348.

[8] Z. Huang, H. Chen, C.J. Hsu, W.H. Chen, S. S. Wu. Credit rating analysis with support vector machines and neural networks: A market comparative study. Decision Support Systems. 37 (4) (2004) 543-558.

[9] A. Khashman. A neural network model for credit risk evaluation. International Journal of Neural Systems. 19 (2009) 285-294.

[10] K. Adnan. Credit risk evaluation using neural networks: Emo-tional versus conventional models. Applied Soft Computing. 11 (8) (2011) 5477-5484.

[11] H.A. Bekhet, S.F.K. Eletter. Credit risk assessment model for Jordanian commercial banks: Neural scoring approach. Review of Development Finance. 4(1) (2014) 20-28.

[12] J. Liu, X. Chen. Credit Risk Assessment Model for Individual Housing Loan Based on Decision Tree. Computer Engineering. 32(13) (2006) 263-265.

[13] C.K. Leong. Credit Risk Scoring with Bayesian Network Models. Computational Economics. 47(3) (2016) 423-446.

[14] H. Zhou, Y. Lan, Y.C. Soh, G.B. Huang. Credit risk evaluation with extreme learning machine. IEEE International Conference on Systems, Man, and Cybernetics. 2(1) (2012) 1064-1069.

[15] J. Chorowski, J. Wang, J.M.Zurada. Review and performance comparison of SVM- and ELM-based classifiers. Neurocomputing. 128 (2014) 507-516.

[16] L. Yu, X Yao, S. Wang, K.K. Lai. Credit risk evaluation using a weighted least squares SVM classifier with design of experiment for parameter selection. Expert Systems with Applications. 38 (12) (2011) 15392-15399.

[17] G. Wang, J. Ma. A hybrid ensemble approach for enterprisecredit risk assessment based on support vector machine. Expert Systems with Applications. 39 (5) (2012) 5325-5331.

[18] T. Harris. Quantitative credit risk assessment using support vector machines: broad versus narrow default definitions. Expert Systems with Applications.40 (2013) 4404-4413.

[19] X. Yao, J. Crook. Galina Andreeva. Support vector regression for loss given default modelling. European Journal of Operational Research. 240 (2015) 528-538.

[20] V.N. Vapnik. The Nature of Statistical Learning Theory. Springer, New York, 1995.

[21] J. Platt. Sequential minimal optimization: A fast algorithm for training support vector machines. Microsoft Research. Technical Report MSR-TR-98-14, 1998.

[22] X.F. Song, W.M. Chen, Y.P.P. Chen, B Jiang. Candidate working set strategybased SMO algorithm in support vector machine, Information Processing and Management. 45 (5) (2009) 584-592.

[23] F. Cai, V. Cherkassky. Generalized SMO Algorithm for SVM-Based Multitask Learning, IEEE Transactions on Neural Networks, vol. 23 (6) (2012) 997-1003.

[24] L.J. Cao, S.S. Keerthi, C.J. Ongb, P. Uvarajc, X.J. Fuc, H.P. Leec. Developing parallel sequential minimal optimization for fast training support vector machine. Neurocomputing. 70 (2006) 93-104.

[25] L.J. Cao, S.S. Keerthi, C.J. Ongb, J.Q. Zhang, H.P. Lee. Parallel sequential minimal optimization for the training of support vector machines, IEEE Transactions on Neural Networks. 17 (4) (2006) 1039-1049.

[26] P.H. Chen. R.E. Fan, C.J. Lin. A study on SMO-type decomposition methods for support vector machines. IEEE Transactions on Neural Networks. 17 (4) (2006) 893-908.

[27] Y.L. Lin, J.G. Hsieh, H.K. Wu, J.H. Jeng. Three-parameter sequential minimal optimization for support vector machines. Neurocomputing. 74 (17) (2011) 3467-3475.

[28] S. Cheng, F.Y. Shih. An improved incremental training algorithm for support vector machines using active query. Pattern Recognition. 40 (3) (2007) 964-971.

[29] E. Osuna, R. Freund, F. Girosi. Improved Training Algorithm for Support Vector Machines. Proc. IEEE NNSP. 97 (1997) 276-285.

[30] X. Zhang, W.G. Zhang, W.J. Xu. An optimization model of the portfolio adjusting problem with fuzzy return and a SMO algorithm. Expert Systems with Applications. 38 (2011) 3069-3074.

[31] S.S. Keerthi, S.K. Shevade, C. Bhattacharyya, K.R.K. Murthy. Improvements to Platt's SMO algorithm for SVM classifier design. Neural Computation. 13 (2001) 637-649.

[32] P. Hajek, K. Michalak. Feature selection in corporate credit rating prediction. Knowledge-Based Systems journal. 51 (2013) 72-84.

[33] J.V. Hulse, T.M. Khoshgoftaar, A. Napolitano, R. Wald. Feature selection with high-dimensional imbalanced data. In: Proceedings of the IEEE International Conference on Data Mining Workshops. (2009) 507-514.

[34] S. Maldonado, R. Weber, F. Famili. Feature selection for high-dimensional class-imbalanced data sets using support vector machines. Information Sciences. 286 (2014) 228-246.

[35] G.G. Sundarkumar, V. Ravi. A novel hybrid undersampling method for mining unbalanced datasets in banking and insurance. Engineering Applications of Artificial Intelligence. 37 (2015) 368-377.

[36] T. Fawcett. An introduction to ROC analysis. Pattern Recognition Letters, 2006, 27: 861-874.

Appendix A

Theorem 1. Suppose $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}$ is a symmetry positive definite matrix, let $x = (x_1, x_2)^T$ and $b = (b_1, b_2)^T$, then the box constrained problem

$$\begin{array}{ll} \min & q(x) = \frac{1}{2}x^T Ax - b^T x \\ s.t & l_i \leq x_i \leq u_i,\ l_i < u_i,\ i = 1, 2 \end{array} \tag{19}$$

has a unique global optimal solution as follows

(I) $a_{12} \geq 0$,

$$\begin{cases} x_1^* = \min(\max(l_1, \overline{x_1}, \frac{b_1 - a_{12}u_2}{a_{11}}), \max(\frac{b_1 - a_{12}l_2}{a_{11}}, l_1), u_1) \\ x_2^* = \min(\max(l_2, \overline{x_2}, \frac{b_2 - a_{12}u_1}{a_{22}}), \max(\frac{b_2 - a_{12}l_1}{a_{22}}, l_2), u_2) \end{cases} \tag{20}$$

(II) $a_{12} < 0$,

$$\begin{cases} x_1^* = \min(\max(l_1, \overline{x_1}, \frac{b_1 - a_{12}l_2}{a_{11}}), \max(\frac{b_1 - a_{12}u_2}{a_{11}}, l_1), u_1) \\ x_2^* = \min(\max(l_2, \overline{x_2}, \frac{b_2 - a_{12}l_1}{a_{22}}), \max(\frac{b_2 - a_{12}u_1}{a_{22}}, l_2), u_2) \end{cases} \tag{21}$$

where $\overline{x_1} = \frac{b_1 a_{22} - b_2 a_{12}}{det(A)}$ and $\overline{x_2} = \frac{-b_1 a_{12} + b_2 a_{11}}{det(A)}$.

Proof: Since problem (19) is strict convex, suppose $x^* = (x_1^*, x_2^*)^T$ is the unique global optimal solution of (19), then we only prove that x^* satisfies the following KKT conditions:

$$\begin{cases} \nabla q(x^*)_i = 0,\ l_i < x_i^* < u_i \\ \nabla q(x^*)_i \geq 0,\ x_i^* = l_i \\ \nabla q(x^*)_i \leq 0,\ x_i^* = u_i,\ i = 1,2 \end{cases}$$

The gradient of function $q(x)$ is

$$\nabla q(x) = \begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

It is easy to verify that $(\overline{x_1}, \overline{x_2})^T$ is the solution of equation $\nabla q(x) = 0$.

Since A is a positive definite matrix, we have $a_{11} > 0,\ a_{22} > 0$ and $det(A) > 0$.

(I) For $a_{12} \geq 0$, there are nine cases discussed as follows:

Case 1: If $l_1 \leq \overline{x_1} \leq u_1,\ l_2 \leq \overline{x_2} \leq u_2$, then $x^* = (\overline{x_1}, \overline{x_2})^T$.

Proof: It is clear that $\nabla q(x^*) = 0$, that is, x^* satisfies KKT condition.

Combining $l_1 \leq \overline{x_1} \leq u_1$ and $l_2 \leq \overline{x_2} \leq u_2$, we have

$$0 < a_{11}\overline{x_1} + a_{12}l_2 \leq b_1 \leq a_{11}\overline{x_1} + a_{12}u_2,$$
$$0 < a_{12}\overline{l_1} + a_{22}\overline{x_2} \leq b_2 \leq a_{12}u_1 + a_{22}\overline{x_2}$$

which implies

$$\frac{b_1 - a_{12}u_2}{a_{11}} \leq \overline{x_1} \leq \frac{b_1 - a_{12}l_2}{a_{11}}, \frac{b_2 - a_{12}u_1}{a_{22}} \leq \overline{x_2} \leq \frac{b_2 - a_{12}l_1}{a_{22}}$$

Therefore $x^* = (\overline{x_1}, \overline{x_2})^T$ is the unique global optimal solution of (19), and it has the form of expression (20).

Case 2: If $\overline{x_1} \geq u_1$ and $\overline{x_2} \geq u_2$, then $x^* = (u_1, u_2)^T$. Proof: By $\overline{x_1} \geq u_1$ and $\overline{x_2} \geq u_2$ we have

$$a_{11}\overline{x_1} + a_{12}l_2 \leq a_{11}\overline{x_1} + a_{12}u_2 \leq b_1$$
$$a_{12}\overline{l_1} + a_{22}\overline{x_2} \leq a_{12}u_1 + a_{22}\overline{x_2} \leq b_2$$

which implies

$$\overline{x_1} \leq \tfrac{b_1 - a_{12}u_2}{a_{11}} \leq \tfrac{b_1 - a_{12}l_2}{a_{11}},$$
$$\overline{x_2} \leq \tfrac{b_2 - a_{12}u_1}{a_{22}} \leq \tfrac{b_2 - a_{12}l_1}{a_{22}},$$
$$\nabla q(x^*) \leq 0$$

namely, $x^* = (u_1, u_2)^T$ satisfies KKT condition and has the form of expression (20). Therefore, x^* is the unique global optimal solution of (19).

Case 3: If $l_1 \leq \overline{x_1} \leq u_1,\ \overline{x_2} > u_2$, then $x^* = (\min(\frac{b_1 - a_{12}u_2}{a_{11}}, u_1), u_2)^T$. Proof: From $l_1 \leq \overline{x_1} \leq u_1,\ \overline{x_2} > u_2$ we get

$$\overline{x_1} \leq \tfrac{b_1 - a_{12}u_2}{a_{11}} \leq \tfrac{b_1 - a_{12}l_2}{a_{11}},$$
$$\tfrac{b_2 - a_{12}u_1}{a_{22}} \leq \overline{x_2} \leq \tfrac{b_2 - a_{12}l_1}{a_{22}},$$

Further we have

$$\begin{cases} \nabla q(x^*)_1 = 0,\ l_1 < \frac{b_1 - a_{12}u_2}{a_{11}} < u_1, \\ \nabla q(x^*)_1 \leq 0, \qquad \frac{b_1 - a_{12}u_2}{a_{11}} \geq u_1, \\ \nabla q(x^*)_2 \leq 0 \end{cases}$$

that is, $x^* = (u_1, u_2)^T$ satisfies KKT condition and has the form of expression (20). Therefore, x^* is the unique global optimal solution of (19).

Case 4: If $\overline{x_1} < l_1$ and $\overline{x_2} > u_2$, then

$$x^* = \begin{cases} (\min(\frac{b_1 - a_{12}u_2}{a_{11}}, u_1), u_2)^T, & \frac{b_1 - a_{12}u_2}{a_{11}} > l_1 \\ (l_1, \min(\max(\frac{b_2 - a_{12}l_1}{a_{22}}, l_2), u_2))^T, & \frac{b_1 - a_{12}u_2}{a_{11}} \leq l_1 \end{cases}$$

Proof: We discuss in two cases.

First, if $\frac{b_1 - a_{12}u_2}{a_{11}} > l_1$, then by $\overline{x_1} < l_1$ and $\overline{x_2} > u_2$ we get $\frac{b_2 - a_{12}l_1}{a_{22}} \geq u_2$ and $x^* = (\min(\frac{b_1 - a_{12}u_2}{a_{11}}, u_1), u_2)^T$. Further we have

$$\begin{cases} \nabla q(x^*)_1 = 0, \quad l_1 < \frac{b_1 - a_{12}u_2}{a_{11}} < u_1, \\ \nabla q(x^*)_1 \leq 0, \qquad \frac{b_1 - a_{12}u_2}{a_{11}} \geq u_1, \\ \nabla q(x^*)_2 \leq 0 \end{cases}$$

that is, $x^* = (\min(\frac{b_1 - a_{12}u_2}{a_{11}}, u_1), u_2)^T$ satisfies KKT condition and has the form of expression (20). Therefore, x^* is the unique global optimal solution of (19).

Second, if $\frac{b_1 - a_{12}u_2}{a_{11}} \leq l_1$, then from $\overline{x_1} < l_1$ and $\overline{x_2} > u_2$ we have $x^* = (l_1, \min(\max(\frac{b_2 - a_{12}l_1}{a_{22}}, l_2), u_2))^T$. Further we get

$$\begin{cases} \nabla q(x^*)_1 \geq 0, \\ \nabla q(x^*)_2 \geq 0, \qquad \frac{b_2 - a_{12}l_1}{a_{22}} \leq l_1, \\ \nabla q(x^*)_2 = 0,\ l_2 < \frac{b_2 - a_{12}l_1}{a_{22}} < u_2, \\ \nabla q(x^*)_2 \leq 0, \qquad \frac{b_2 - a_{12}l_1}{a_{22}} \geq u_2, \end{cases}$$

that is, $x^* = (l_1, \min(\max(\frac{b_2 - a_{12}l_1}{a_{22}}, l_2), u_2))^T$ satisfies KKT condition and has the form of expression (20). Thus, x^* is the unique global optimal solution of (19).

Case 5: If $\overline{x_1} < l_1$ and $l_2 \leq \overline{x_2} \leq u_2$, then $x^* = (l_1, \max(\frac{b_2 - a_{12}l_1}{a_{22}}, l_2))^T$.

The proof is similar to Case 3.

Case 6: If $\overline{x_1} < l_1$ and $\overline{x_2} < l_2$, then $x^* = (l_1, l_2)^T$.

Proof: From $\overline{x_1} < l_1$ and $\overline{x_2} < l_2$ we have

$$\overline{x_1} \geq \frac{b_1 - a_{12} l_2}{a_{11}} \geq \frac{b_1 - a_{12} u_2}{a_{11}}$$
$$\overline{x_2} \geq \frac{b_2 - a_{12} l_1}{a_{22}} \geq \frac{b_2 - a_{12} u_1}{a_{22}}$$

Further we get $\nabla q(x^*) \geq 0$, that is $x^* = (l_1, l_2)^T$ satisfies KKT condition and has the form of expression (20). Therefore, x^* is the unique global optimal solution of (19).

Case 7: If $l_1 \leq \overline{x_1} \leq u_1$ and $\overline{x_2} < l_2$, then $x^* = (\max(\frac{b_2 - a_{12} l_2}{a_{11}}, l_1), l_2)^T$. The proof is similar to Case 3.

Case 8: If $\overline{x_1} < l_1$ and $\overline{x_2} > u_2$, then

$$x^* = \begin{cases} (\max(\frac{b_2 - a_{12} l_2}{a_{11}}, l_1), l_2)^T, & \frac{b_1 - a_{12} l_2}{a_{11}} > u_1 \\ (u_1, \min(\max(\frac{b_2 - a_{12} u_1}{a_{22}}, l_2), u_2))^T, & \frac{b_1 - a_{12} u_2}{a_{11}} \geq u_1 \end{cases}$$

The proof is similar to Case 4.

Case 9: If $\overline{x_1} > u_1$ and $l_2 \leq \overline{x_2} \leq u_2$, then $x^* = (\max(\frac{b_2 - a_{12} l_2}{a_{11}}, l_1), l_2)^T$. The proof is similar to Case 3.

(II) For $a_{12} < 0$, firstly we prove that for any real number a, b, c , the conclusion as follows is right.

(i) $\max[\min(a, c), \min(b, c)] = \min[\max(a, b), c]$

(ii) $\max[\min(a, b), c] = \min[\max(a, c), \max(b, c)]$.

Proof: (i) If $a \leq b$,then $\min(a, c) \leq \min(b, c)$, it is clear that the left and right of the conclusion (i) are equal to $\min(b, c)$; if $b < a$, then the left and right of the conclusion (i) are equal to $\min(a, c)$, thus the conclusion (i) is right; (ii) The proof is similar to (i).

Secondly, substituting $x_1 = y_1$, $x_2 = -y_2$ into the problem (19), we have

$$\min q(x) = p(y) = \frac{1}{2}(y_1\ y_2)\begin{pmatrix} a_{11} & -a_{12} \\ -a_{12} & a_{22} \end{pmatrix}(y_1\ y_2)^T - (b_1 - b_2)(y_1\ y_2)^T$$

$$s.t\ l_1 \leq y_1 \leq u_1,$$
$$-u_2 \leq y_2 \leq -l_2,\ l_i < u_i,\ i = 1, 2 \tag{22}$$

It is clear that $\begin{pmatrix} a_{11} & -a_{12} \\ -a_{12} & a_{22} \end{pmatrix}$ is a positive definite matrix and $-a_{12} > 0$. on the basis of (I), we get the optimal solution of (22) as follows:

$$y_1^* = \min\left(\max\left(l_1, \overline{y_1}, \frac{b_1 - a_{12} l_2}{a_{11}}\right), \max\left(\frac{b_1 - a_{12} u_2}{a_{11}}, l_1\right), u_1\right)$$

$$y_2^* = \min\left(\max\left(-u_2, \overline{y_2}, \frac{-b_2 + a_{12} u_1}{a_{22}}\right), \max\left(\frac{-b_2 + a_{12} l_1}{a_{22}}, -u_2\right), -l_2\right)$$

where $\overline{y_1} = \frac{b_1 a_{22} - b_2 a_{12}}{det(A)} = \overline{x_1}$, $\overline{y_2} = \frac{b_1 a_{12} - b_2 a_{11}}{det(A)} = -\overline{x_2}$. Finally, we derive the optimal solution of (19) as follows:

$$x_1^* = y_1^* = \min\left(\max\left(l_1, \overline{x_1}, \frac{b_1 - a_{12} l_2}{a_{11}}\right), \max\left(\frac{b_1 - a_{12} u_2}{a_{11}}, l_1\right), u_1\right)$$

$$x_2^* = -y_2^*$$
$$= -\min\left(\max\left(-u_2, -\overline{x_2}, -\frac{b_2 - a_{12} u_1}{a_{22}}\right), \max\left(-\frac{b_2 - a_{12} l_1}{a_{22}}, -u_2\right), -l_2\right)$$

$$= \max\left[\min\left(u_2, \overline{x_2}, \frac{b_2 - a_{12} u_1}{a_{22}}\right), \min\left(\frac{b_2 - a_{12} l_1}{a_{22}}, u_2\right), l_2\right]$$

$$= \max\left\{\max\left[\min\left(u_2, \min\left(\overline{x_2}, \frac{b_2 - a_{12} u_1}{a_{22}}\right)\right), \min\left(\frac{b_2 - a_{12} l_1}{a_{22}}, u_2\right)\right], l_2\right\}$$

$$\underline{\underline{(i)}} \max\left\{\min\left[\max\left(\min\left(\overline{x_2}, \frac{b_2 - a_{12} u_1}{a_{22}}\right), \frac{b_2 - a_{12} l_1}{a_{22}}\right), u_2\right], l_2\right\}$$

$$= \max\left\{\min\left[\min\left(\max\left(\overline{x_2}, \frac{b_2 - a_{12} l_1}{a_{22}}\right), \frac{b_2 - a_{12} u_1}{a_{22}}\right), u_2\right], l_2\right\}$$

$$= \max\left\{\min\left[\max\left(\overline{x_2}, \frac{b_2 - a_{12} l_1}{a_{22}}\right), \frac{b_2 - a_{12} u_1}{a_{22}}, u_2\right], l_2\right\}$$

$$= \max\left\{\min\left[\max\left(\overline{x_2}, \frac{b_2 - a_{12} l_1}{a_{22}}\right), \min\left(\frac{b_2 - a_{12} u_1}{a_{22}}, u_2\right)\right], l_2\right\}$$

$$\underline{\underline{(ii)}} \min\left\{\max\left[\max\left(\overline{x_2}, \frac{b_2 - a_{12} l_1}{a_{22}}\right), l_2\right], \max\left[\min\left(\frac{b_2 - a_{12} u_1}{a_{22}}, u_2\right), l_2\right]\right\}$$

$$= \min\left\{\max\left(\overline{x_2}, \frac{b_2 - a_{12} l_1}{a_{22}}, l_2\right), \min\left[\max\left(\frac{b_2 - a_{12} u_1}{a_{22}}, l_2\right), u_2\right]\right\}$$

$$= \min\left(\max\left(\overline{x_2}, \frac{b_2 - a_{12} l_1}{a_{22}}, l_2\right), \max\left(\frac{b_2 - a_{12} u_1}{a_{22}}, l_2\right), u_2\right)$$

In summary, x^* is the optimal solution of (19) and has the form of expression (21).

qi zhang Qi Zhang is currently a PhD student in the University of Chinese Academy of Sciences, Beijing, China. She is also a PhD student of the joint program between City University of Hong Kong and University of Chinese Academy of Sciences. Her current research interests include computational intelligence, research social network and scientometric analysis.

Jue Wang Associate Professor of Chinese Academy of Sciences. Research field concludes computational intelligence, decision analysis, economic forecasting. More than 70 papers had been published in "European Journal of Operational Research", "Fuzzy Optimization and Decision Making", "Experts System with Application" and so on. 6 books are published in America IGI Global publisher, and The Scientific Publisher in China. As executive editors of "Economic Forecasting Science Series", "low carbon economy and Chinese development" etc. and editorial board member of many important international journals. Hosted and participated in a number of National Natural Science Fund project, the knowledge innovation project of Chinese Academy of Sciences, Ministry of science and technology project assessment project, the financial risk management project, large Petrochemical Industries.

Aiguo Lu She was born in Yicheng, ShanXi Province, China. She received her Ph.D. degree in Department of Mathematics from Xidian University. Her research interests include Support vector machines and Unconstrained Optimization Problems, and so on.

Shouyang Wang received the Ph.D. degree in operations research from Institute of Systems Science, Chinese Academy of Sciences (CAS), Beijing in 1986. He is currently a Bairen distinguished professor of Management Science at Academy of Mathematics and Systems Sciences of CAS and a Lotus chair professor of Hunan University, Changsha. He is the editor-in-chief or a co-editor of 12 journals. He has published 18 books and over 120 journal papers. His current research interests include financial engineering, e-auctions and decision support systems.

Jian Ma is Professor in Department of Information Systems at City University of Hong Kong. He is specialised in the research areas of research information systems, business intelligence and research social networks. He has publised over 120 journal articles with SCI H index 22. His applied research work has been widely used in government funding agencies (e.g. National Natural Science Foundation of China) and universities (e.g. University of Hong Kong.