

# An Improved Fuzzy Support Vector Machine for Credit Rating

Yanyou Hao<sup>1,2</sup>, Zhongxian Chi<sup>1</sup>, Deqin Yan<sup>3</sup>, Xun Yue<sup>1</sup>

<sup>1</sup> Department of Computer Science and Engineering,  
Dalian University of Technology, Dalian, 116024, China

<sup>2</sup> Dalian Branch of CCB, Dalian, 116001, China

<sup>3</sup>Department of Computer Science, Liaoning Normal University,  
Dalian, 116029, China  
haoyanyou@gmail.com

**Abstract.** In order to classify data with noises or outliers, Fuzzy support vector machine (FSVM) improve the generalization power of traditional SVM by assigning a fuzzy membership to each input data point. In this paper, an improved FSVM based on vague sets is proposed by assigning a truth-membership and a false-membership to each data point. And we reformulate the improved FSVM so that different input points can make different contributions to decision hyperplane. The effectiveness of the improved FSVM is verified in credit rating; the experiment results show that our method is promising.

**Keywords:** Fuzzy support vector machine (FSVM); fuzzy membership; vague sets; credit rating.

## 1 Introduction

Support vector machine (SVM) is based on the statistical learning theory (SLT) developed by Vapnik [1]. The formulations of SVM embody the structural risk minimization (SRM) principle and VC dimensions theory. Due to their high generalization ability SVMs have attracted many attentions and have been accepted widely [2, 3]. Credit risk is the primary risk facing commercial banks. With the proposed guidelines under the New Basel Accord, financial institutions will benefit from better assessing their risks. Statistical methods have been used to estimate credit rating, which typically require large data to build the forecasting model. However, there are not large data to use in real-life. Huang and Wang have shown that SVMs achieved better performance than traditional statistical methods in the field of credit rating assessment [4] [8].

The aim of SVM is to provide good generalization ability. The optimal hyperplane can be determined by only few data points that are called support vectors (SVs). Accordingly, SVMs can provide a good generalization performance for classification problems despite it does not incorporate problem-domain knowledge. There are some advantages of SVMs as follows. The training of SVMs is relatively easy and has no

local optimal like neural networks. And SVMs scales relatively well to high dimensional data and tradeoff between classifier complexity and error. But there are some disadvantages in SVMs which need a “good” kernel function and the overfitting problem like neural network. These disadvantages impact the generalization ability of SVMs in many real applications.

As remarked in [5, 6, 7], SVMs are very sensitive to outliers and noises. The FSVMs proposed in [5, 6] treat the training data points with different importance in the training process. Namely, FSVMs fuzzily the penalty term of the cost function to be minimized, reformulate the constrained optimization problem, and then construct the Lagrangian so that the solutions to the optimal hyperplane in the primal form can be found in the dual form. The key part of FSVM is how to construct the membership model in the training data set. However, there is no general rule to determine the membership of each data point now. Wang proposed the Bilateral-weight fuzzy SVM (B-FSVM) to evaluate the credit risk [8]. This method treats each instance as both of positive and negative classes, but assigned with different memberships.

In this paper, we propose an improved fuzzy SVM based on the vague sets. We assign a truth-membership and a false-membership for each data point in training data set, and reformulate the FSVM. The experiment results show this approach improve the generalization performance of traditional SVM.

The rest of this paper is organized as follows. A brief review of the theory of FSVMs, B-FSVMs and Vague Sets will be given in Sections 2. Section 3 presents an improved FSVM based on vague sets. Section 4 reports the experiment results and analysis. Section 5 presents conclusions and future work.

## **2 Fuzzy Support Vector Machines and Vague Sets**

### **2.1 Fuzzy Support Vector Machines**

SVM has some merits such as multi-local minima and overfitting in neural networks seldom occur in SVM and it has a solid theoretical foundation. However, there are still some problems in SVM. All training points of one class are treated uniformly in the theory of SVM. As shown in [5,6], due to overfitting, the training process is very sensitive to those outliers in the training dataset which are far away from their own class.

FSVM is proposed in order to decrease the effect of those outliers. The main idea of FSVM is that we can assign a fuzzy membership to each data point, In other words, FSVM treats the input data unequally such that different data points can have different effects in the learning of the separating hyperplane. We can treat the noises or outliers as less importance and let these points have lower fuzzy membership. It is also based on the maximization of the margin like the traditional SVM, but FSVM uses fuzzy memberships to prevent some points from making narrower margin. This equips

FSVM with the ability to train data with outliers by setting lower fuzzy memberships to the data points that are considered as noises or outliers with higher probability [9].

We give a briefly introduction about FSVM that are proposed in [5, 6] as follows. Suppose we are given a set of labeled training data sets with associated fuzzy membership  $(y_i, x_i, m_i)$ . Each training point  $x_i \in R^N$  is given a label  $y_i \in \{-1, 1\}$  and a fuzzy membership  $\varepsilon \leq m_i \leq 1$  with  $i = 1, 2, \dots, N$ , and sufficient small  $\varepsilon > 0$ . Since the fuzzy membership  $m_i$  is the attitude of the corresponding point  $x_i$  toward one class and the parameter  $\xi_i$  is a measure of error in SVM, the term  $m_i \xi_i$  is a measure of error with different weighting. Then the optimal hyperplane problem is regarded as the solution to

$$\min \frac{1}{2} w^T \cdot w + C \sum_{i=1}^N m_i \xi_i, \quad (1)$$

$$\text{Subject to: } \begin{cases} y_i(w \cdot x_i + b) \geq 1 - \xi_i \\ i = 1, \dots, N; \quad \xi_i \geq 0 \end{cases}.$$

Where  $C$  is a free parameter that controls the tradeoff between the maximum of margin and minimum of classification error. It is noted that a smaller  $m_i$  reduces the effect of the parameter  $\xi_i$  in problem (1) such that the corresponding point is treated as less important. We obtain the following decision function

$$f(x) = \text{sign}(w \cdot x + b) = \text{sign} \left( \sum_{i=1}^N \alpha_i y_i K(x_i \cdot x) + b \right). \quad (2)$$

$$0 \leq \alpha_i \leq m_i C. \quad (3)$$

Wang proposed the Bilateral-weight fuzzy SVM (B-FSVM) and presented a new approach to improve the FSVM [2]. This new method treats every data point in the training dataset as both positive and negative class but with different memberships. Memberships are assigned to both classes for every data point. This means it increases the number of training data points from the original  $N$  to  $2*N$ . i.e. from  $\{x_i, y_i\}$  to  $\{x_i, 1, m_i\}, \{x_i, -1, 1 - m_i\}$ . The classification problem is modeled by the following programming

$$\min \frac{1}{2} w^T w + C \sum_{i=1}^N [m_i \xi_i + (1 - m_i) \eta_i], \quad (4)$$

$$\text{Subject to: } \begin{cases} [w^T \phi(x_i) + b] \geq 1 - \xi_i, i = 1, \dots, N \\ [w^T \phi(x_i) + b] \leq -1 + \eta_i, i = 1, \dots, N \\ \xi_i \geq 0, \text{ for } i = 1, \dots, N \\ \eta_i \geq 0, \text{ for } i = 1, \dots, N. \end{cases}$$

The problem can be transformed into its dual form

$$\max_{\alpha_i, \beta_i} \sum_{i=1}^N \alpha_i + \sum_{i=1}^N \beta_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i - \beta_i)(\alpha_j - \beta_j) \phi(x_i)^T \phi(x_j) , \quad (5)$$

$$\sum_{i=1}^N \alpha_i = \sum_{i=1}^N \beta_i ,$$

$$\text{Subject to: } 0 \leq \alpha_i \leq C m_i, i = 1, \dots, N, \\ 0 \leq \beta_i \leq C(1 - m_i), i = 1, \dots, N.$$

Let  $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ . In order to transform this into a quadratic programming problem, we let  $\gamma_i = \alpha_i - \beta_i$ . The previous optimization becomes

$$\max_{\beta_i, \gamma_i} \sum_{i=1}^N \gamma_i + \sum_{i=1}^N 2\beta_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \gamma_i \gamma_j K(x_i, x_j) , \quad (6)$$

$$\sum_{i=1}^N \gamma_i = 0.$$

$$\text{Subject to: } 0 \leq \beta_i + \gamma_i \leq C m_i \text{ for } i = 1, \dots, N. \\ 0 \leq \beta_i \leq C(1 - m_i) \text{ for } i = 1, \dots, N.$$

After solving this quadratic programming problem, we obtain the following classifier

$$f(x) = \text{sign}(w^T \phi(x) + b) = \text{sign}\left(\sum_{i=1}^N (\alpha_i - \beta_i) K(x, x_i) + b\right). \quad (7)$$

## 2.2 Vague Sets

Since the theory of fuzzy sets was proposed in 1965 by Zadeh [10], it has been used for handling fuzzy decision-making problems and the study of the aggregation processes.

**Definition 1. (Fuzzy Set)** Let  $U$  be the universe of discourse,  $U = \{x_1, x_2, \dots, x_n\}$ . A fuzzy set

$$A = \{(x, \mu_A(x)) \mid x \in A, \mu_A(x) \in [0, 1]\}$$

The grade of membership of an element  $x_i$  in a fuzzy set is denoted as  $\mu_A(x_i)$ ,  $x_i \in U$ . It is represented by a real value between zero and one. Fuzzy sets assign each object a single value. This single value combines the evidence for  $x_i \in U$  and the evidence against  $x_i \in U$ , without indicating how much there is of each. The single number tells us nothing about its accuracy.

Vague set theory, introduced by Gau and Buehrer in 1993 [11], extends and improves fuzzy set theory. The membership ranging between zero and one in fuzzy set theory is extended to a continuous subinterval of  $[0, 1]$ .

**Definition 2. (Vague Set)** Let  $U$  be the universe of discourse, with a generic element  $x_i \in U$ . A vague set  $V$  in  $U$  is characterized by a truth membership function  $t_V$  and a false membership function  $f_V$ ,  $t_V(x_i)$  is a lower bound on the grade of membership of  $x_i$  derived from the evidence for  $x_i$ , and  $f_V(x_i)$  is a lower bound on the negation of  $x_i$  derived from the evidence against  $x_i$ .  $t_V(x_i)$  and  $f_V(x_i)$  both associate a real number in the interval  $[0,1]$  with each point in  $U$ , where  $t_V(x_i) + f_V(x_i) \leq 1$ .

This approach bounds the grade of membership of  $x_i$  to a subinterval  $[t_V(x_i), 1 - f_V(x_i)]$  of  $[0, 1]$ . The vague value can be divided into three parts: the truth-membership part  $t_V(x_i)$ , the false-membership part  $f_V(x_i)$ , and the unknown part  $1 - t_V(x_i) - f_V(x_i)$ .

### 3 FSVM Based on Vague Sets (VS-FSVM)

A vague set is a further generalization of a fuzzy set. Instead of using point-based membership as in fuzzy sets, interval-based membership is used in a vague set [12]. The interval-based membership in vague sets is more expressive in capturing uncertainty and vagueness of data. In this paper, we use vague sets to define the membership in FSVMs.

Given a training datasets  $\{x_i, y_i\}$ , for  $i = 1, \dots, N$ ,  $y_i \in \{-1, 1\}$ , we transform the original data sets into a new training datasets

$$\{x_i, 1, t_i\}, \{x_i, -1, f_i\}, \text{ for } i = 1, \dots, N.$$

$t_i$  is the truth-membership of each data point in the training datasets with positive class.

$f_i$  is the false-membership of each data point in the training datasets with negative class.

$$t_i \leq 1, f_i \leq 1 \text{ and } t_i + f_i \leq 1.$$

Let  $m_i = 1 - t_i - f_i$ , which is the unknown part indicating the probability of each data point without positive class or negative class in the training datasets.

We propose an improved fuzzy SVM based on vague sets and deduce the formulation as follows.

$$\min_{w, b, \xi, \eta, \theta} \frac{1}{2} w^T w + C \sum_{k=1}^N [t_i \xi_i + f_i \eta_i + m_i \theta_i], \quad (8)$$

$$w^T x_i + b \geq 1 - \xi_i - \lambda \theta_i, \text{ for } i = 1, \dots, N$$

$$w^T x_i + b \leq -1 + \eta_i + (1 - \lambda) \theta_i, \text{ for } i = 1, \dots, N$$

$$\text{Subject to: } \xi_i \geq 0, \text{ for } i = 1, \dots, N$$

$$\eta_i \geq 0, \text{ for } i = 1, \dots, N$$

$$\theta_i \geq 0, \text{ for } i = 1, \dots, N.$$

Where  $\theta_i$  is a slack variable of  $m_i$ ,  $\lambda$  is a free parameter less than 1 which segments the unknown part, we process it with exponential value. To solve the previous optimization problem, letting the corresponding Lagrange multipliers to the condition be  $\alpha_i, \beta_i, \mu_i, \nu_i$  and  $\tau_i$ , we construct the Lagrangian function

$$\begin{aligned} L(w, b, \xi, \eta, \theta) &= \frac{1}{2} w^T w + C \sum_{i=1}^N [t_i \xi_i + f_i \eta_i + m_i \theta_i] \\ &\quad - \sum_{i=1}^N \alpha_i (w^T x_i + b - 1 + \xi_i + \lambda \theta_i) + \sum_{i=1}^N \beta_i (w^T x_i + b + 1 - \eta_i - (1 - \lambda) \theta_i) \\ &\quad - \sum_{i=1}^N \mu_i \xi_i - \sum_{i=1}^N \nu_i \eta_i - \sum_{i=1}^N \tau_i \theta_i \end{aligned} \quad (9)$$

And we find the saddle point of  $L$ , where  $\alpha_i \geq 0, \beta_i \geq 0, \mu_i \geq 0, \nu_i \geq 0$  and  $\tau_i \geq 0$ . These parameters must satisfy the following conditions

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^N \alpha_i x_i + \sum_{i=1}^N \beta_i x_i = 0. \quad (10)$$

$$\frac{\partial L}{\partial b} = -\sum_{i=1}^N \alpha_i + \sum_{i=1}^N \beta_i = 0. \quad (11)$$

$$\frac{\partial L}{\partial \xi_i} = C t_i - \alpha_i + \mu_i = 0. \quad (12)$$

$$\frac{\partial L}{\partial \eta_i} = C f_i + \beta_i + \nu_i = 0. \quad (13)$$

$$\frac{\partial L}{\partial \theta_i} = C u_i - \lambda \alpha_i + (1 - \lambda) \beta_i - \tau_i = 0. \quad (14)$$

From the Kuhn-Tucker Theorem, the following conditions are also satisfied

$$\alpha_i (w^T x_i + b - 1 + \xi_i + \lambda \theta_i) = 0. \quad (15)$$

$$\beta_i (w^T x_i + b + 1 - \eta_i - (1 - \lambda) \theta_i) = 0. \quad (16)$$

$$\mu_i \xi_i = 0, \nu_i \eta_i = 0, \tau_i \theta_i = 0. \quad (17)$$

$$\begin{aligned} \xi_i \geq 0, \eta_i \geq 0, \theta_i \geq 0, \alpha_i \geq 0, \beta_i \geq 0, \\ \tau_i \geq 0, \mu_i \geq 0, \nu_i \geq 0, i = 1, \dots, N. \end{aligned} \quad (18)$$

According to (10), we obtain

$$w = \sum_{i=1}^N (\alpha_i - \beta_i) x_i. \quad (19)$$

We define  $E$  and  $F$  as

$$E = -\sum_{i=1}^N \alpha_i (w^T x_i + b - 1 + \xi_i + \lambda \theta_i). \quad (20)$$

$$F = \sum_{i=1}^N \beta_i (w^T x_i + b + 1 - \eta_i - (1 - \lambda) \theta_i). \quad (21)$$

Apply  $E$  and  $F$  into (9)

$$L = \frac{1}{2} w^T w + C \sum_{i=1}^N [t_i \xi_i + f_i \eta_i + m_i \theta_i] + E + F - \sum_{i=1}^N \mu_i \xi_i - \sum_{i=1}^N \nu_i \eta_i - \sum_{i=1}^N \tau_i \theta_i. \quad (22)$$

From (15) and (16), we get  $E=F=0$ , and (22) be transformed into

$$\begin{aligned} L &= \frac{1}{2} w^T w + \sum_{i=1}^N (C t_i - \mu_i) \xi_i + \sum_{i=1}^N (C f_i - \nu_i) \eta_i + \sum_{i=1}^N (C m_i - \tau_i) \theta_i \\ &= \frac{1}{2} w^T w + \sum_{i=1}^N (\alpha_i \xi_i + \beta_i \eta_i) + \sum_{i=1}^N [\lambda \alpha_i + (1 - \lambda) \beta_i] \theta_i. \end{aligned} \quad (23)$$

Taking the difference between the sum of (15) and (16) from  $i=1$  to  $N$ , we obtain

$$\sum_{i=1}^N (\alpha_i \xi_i + \beta_i \eta_i) + \sum_{i=1}^N [\lambda \alpha_i + (1 - \lambda) \beta_i] \theta_i = -\sum_{i=1}^N (\alpha_i - \beta_i) w^T x_i + \sum_{i=1}^N (\alpha_i + \beta_i). \quad (24)$$

Apply (24) into (23), we obtain

$$L = \frac{1}{2} w^T w - \sum_{i=1}^N (\alpha_i - \beta_i) w^T x_i + \sum_{i=1}^N (\alpha_i + \beta_i). \quad (25)$$

Apply (19) into (25), we obtain

$$L = \sum_{i=1}^N (\alpha_i + \beta_i) - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i - \beta_i) (\alpha_j - \beta_j) x_i^T x_j. \quad (26)$$

The original classification problem can be transformed into its dual form

$$\max \sum_{i=1}^N (\alpha_i + \beta_i) - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i - \beta_i) (\alpha_j - \beta_j) x_i^T x_j, \quad (27)$$

$$\sum_{i=1}^N \alpha_i = \sum_{i=1}^N \beta_i$$

$$\text{Subject to: } 0 \leq \alpha_i \leq C t_i + \lambda m_i, \quad i = 1, \dots, N.$$

$$0 \leq \beta_i \leq C f_i + (1 - \lambda) m_i, \quad i = 1, \dots, N.$$

Let  $K(x_i, x_j) = x_i^T x_j$ , in order to transform (27) into a quadratic programming problem, we let  $\gamma_i = \alpha_i - \beta_i$ . The previous optimization becomes

$$\max \sum_{i=1}^N \gamma_i + 2 \sum_{i=1}^N \beta_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \gamma_i \gamma_j K(x_i, x_j), \quad (28)$$

$$\sum_{i=1}^N \gamma_i = 0$$

$$\text{Subject to: } 0 \leq \alpha_i \leq Ct_i, i = 1, \dots, N$$

$$0 \leq \beta_i \leq Cf_i, i = 1, \dots, N$$

$$0 \leq \lambda \alpha_i + (1 - \lambda) \beta_i \leq Cm_i, i = 1, \dots, N$$

After solving this and substituting  $w = \sum_{i=1}^N (\alpha_i - \beta_i) \phi(x_i)$  into the original classification problem, we obtain the following classifier:

$$y(x_i) = \text{sign}(w^T \phi(x_i) + b) = \text{sign} \left( \sum_{i=1}^N (\alpha_i - \beta_i) K(x \cdot x_i) + b \right). \quad (29)$$

## 4 Experiments Results and Analysis

### 4.1 Data Sets and Criteria

We used real life home loan data to conduct the experiment and selected one thousand sample data from a major commercial bank of China. The attributes of these data consisted of the customer information and the other information of the loan application form. There are 16 attributes listed in Table 1.

**Table 1.** Samples' Attributes

Index	Attributes
A01	Year of birth
A02	Number of children
A03	Number of other dependents
A04	Is there a home phone
A05	Education Level
A06	Applicant's Occupation
A07	Spouse's income
A08	Applicant's income
A09	Applicant's employment status
A10	Residential status
A11	Value of home
A12	Mortgage balance outstanding
A13	Monthly repayment of mortgage
A14	Year of Mortgage
A15	Outgoings on loans
A16	Outgoings on credit cards

In order to construct a two-class classification task, “good” customer and “bad” customer were taken into consideration. The “good” customer is the borrower who repayment on time. The “bad” customer is the one that did not pay one instalment over a period of three months. The “good” customer is labeled “1” and the “bad” customer is labeled “-1”. These data is typically from a sample of applicants who have been granted bank credit already.

The estimation of loan defaults is a two-class classification task. Accuracy is the typical performance measure for two-class classification schemes. However, two learning algorithms can have the same accuracy, but the one which groups the errors near the decision border is the better one.

In order to appraise the performance of the classifier, Default Accuracy and Normal Accuracy are selected as standard criteria. We define them as follows

$$\text{Default Accuracy} = \frac{\text{default samples classified}}{\text{total default samples}} \quad (30)$$

$$\text{Normal Accuracy} = \frac{\text{normal samples classified}}{\text{total normal samples}} \quad (31)$$

The advantage of the default accuracy and normal accuracy is that they are a good indicator of whether the errors are close to the decision border or not. Given two classifiers with the same accuracy, the one with high default accuracy and normal accuracy is the better one. This definition is equivalent to the definitions of False Alarm and Miss Rate in [13].

## 4.2 Generating Vague Memberships

The key step of our experiment is to generate vague membership from the result of the loan classification system which is adopted in Chinese commercial banks widely. The loan classification system which is a risk-based approach provides a five-classification management on loan quality. The loans are classified into five categories, that is, pass, special-mention, substandard, doubtful and loss, the last three categories are recognized as non-performing loans (NPLs). The definitions of the five categories are as follow [14].

Pass: borrowers can honor the terms of the contracts, and there is no reason to doubt their ability to repay principal and interest of loans in full and on a timely basis.

Special-mention: borrowers are still able to service the loans currently, although the repayment of loans might be adversely affected by some factors.

Substandard: borrowers' ability to service loans is apparently in question, cannot depend on their normal business revenues to pay back the principal and interest of loans and certain losses might incur even when guarantees are executed.

Doubtful: borrowers cannot pay back principal and interest of loans in full and significant losses will incur even when guarantees are executed.

Loss: principal and interest of loans cannot be recovered or only a small portion can be recovered after taking all possible measures and resorting to necessary legal procedures.

The five-category classification of loan records all status in the management process of loans. The special-mention category is a transitional status between performing loan and non-performing loan. We use the results of five-category classification of loan to generate the vague membership. Then the truth-membership of a borrower is the pass category (performing loan) in proportion to all his loans classifications. The false-membership of a borrower is the non-performing loan category (i.e. substandard, doubtful and loss) in proportion to all his loans classifications. And the special-mention category indicates the unknown part.

### 4.3 Experiment Results

To decrease the bias due to the choice of split between training datasets and test datasets, we randomly divide the original data into two sub-datasets evenly, one half for training and the other half for testing, this is referred to as 50-50 split experiment. The training dataset includes 500 samples with 120 “bad” customers and 380 “good” customers.

We compare our VS-FSVM with the traditional SVM, FSVM and B-FSVM on the same training dataset. Then we predict the same test dataset using their classification model respectively. Table 2 shows the comparison of the predicting result of these algorithms.

**Table 2.** The comparison of predicting result

<b>Classification Algorithm</b>	<b>Normal Accuracy</b>	<b>Default Accuracy</b>	<b>Overall Accuracy</b>
SVM	88.95%	69.17%	84.20%
FSVM	89.47%	70.83%	85.00%
B-FSVM	90.26%	72.50%	86.00%
VS-FSVM	91.84%	75.00%	87.80%

It is obvious in Table 2 that total accuracy of VS-SVM is better than other SVMs, that is, the predicting accuracy can increase just 2%. Our future direction of the research will focus on how to improve the Default Accuracy especially in the test dataset.

## 5 Conclusions

In this paper, we propose an improved FSVM based on vague sets by assigning a truth-membership and a false-membership to each data point. And we reformulate the improved FSVM so that different input points can make different contributions to decision hyperplane. In order to verify the performance of the improved FSVM, we

apply it to credit assessment. The experiment results show that our method is promising. Our future direction of the research will focus on how to improve the default accuracy. We believe that deeper data preprocessing and more suitable parameters selection will improve the performance of generalization. Extending the two-class classification to multi-class classification is also our future research work.

## References

1. V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.
2. C. Cortes and V. N. Vapnik, "Support vector networks," *Machine Learning*, vol. 20, pp. 273- 297, 1995.
3. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, 1998.
4. Zan Huang and Hsinchun Chen.: Credit rating analysis with support vector machines and neural networks: a market comparative study. *Decision Support Systems* 37(2004) 543-558.
5. Chun-Fu Lin, Sheng-De Wang, *Fuzzy Support Vector Machines*. *IEEE Transactions on Neural Networks*. 2 (2002) 464- 471.
6. Huang, H.P., Liu, Y.H.: *Fuzzy Support Vector Machines for Pattern Recognition and Data Mining*, *International Journal of Fuzzy Systems*. 4 (2002) 826- 835.
7. X. Zhang, "Using class-center vectors to build support vector machines," in *Proc. IEEE NNSP'99*, 1999, pp. 3-11.
8. Yong-Qiao Wang, Shou-Yang Wang and K.K. Lai, A New Fuzzy Support Vector Machine to Evaluate Credit Risk, *IEEE Trans. On Fuzzy System*, Vol. 13, No. 6, Dec 2005.
9. Chun-Fu Lin, Sheng-De Wang, Training algorithms for fuzzy support vector machines with noisy data, *IEEE XIII Workshop on Neural Networks for Signal Processing*, p517-p526, 2003.
10. Zadeh, L. A. Fuzzy sets. *Information and Control*, 1965, 8 (3): 338- 353.
11. Wen-Lung Gau, Daniel J. Buehrer. Vague sets. *IEEE Transactions on Systems, Man and Cybernetics*, 1993, 23 (2) : 610- 614.
12. An Lu, Wilfred Ng, *Vague Sets or Intuitionistic Fuzzy Sets for Handling Vague Data: Which One Is Better?* *ER* 2005: 401-416.
13. Drucker H., Wu D., V. Vapnik.: *Support Vector Machines for Spam Categorization*. *IEEE Transactions on Neural Networks*. Vol. 10, Issue 5, Sep 1999.
14. The People's Bank of China, *Guidelines on Risk-Based Loan Classification*, 2001, <http://www.pbc.gov.cn/english/detail.asp?col=6800&ID=46>.