A Strategy of Maximizing the Sum of Weighted Margins for Ranking Multi Classification Problem

Linkai Luo[1,2]
[1]Department of Automation
Xiamen University
Xiamen, P.R.China,361005

Chengde Lin[1,*],Hong Peng[1],Qifeng Zhou[1]
[2]Research Center of System and Control
Xiamen University
Xiamen, P.R.China, 361005
[*]Corresponding Author: cdlin@xmu.edu.cn

***Abstract*—This paper discusses the strategies of maximizing the sum of margins for ranking multi classification problem. First, the strategy of maximizing the sum of margins (MSM) is extended to maximizing the sum of weighted margins (MSWM). Using MSWM, a mathematical model is established to deal with the ranking multi classification problems where the importance of margins between classes is different, and its dual model is deduced. Then, by introducing the concept of algebraic margin, which is a generalization of geometric margin, the MSWM is further extended to maximizing the sum of weighed algebraic margins (MSWAM). Based on the MSWAM, the deduced mathematical model of the ranking multi classification problem not only has positive generalization ability, but is also a simple linear programming model.**

***Keywords*—ranking multi classification,SVM,sum of weighted margin, algebraic margin**

I. INTRODUCTION

Support Vector Machine (SVM) is originally designed to solve binary classification problem. How to extend it efficiently to deal with multi classifying problem is a very important research hotspot. In general, there are three types of methods to solve the multi classifying problem based on SVM [1]. The first one is the combination method, which combines several binary classifiers (SVMs) under certain strategies, such as one-against-all approach [2], one-against-one approach [3] and DAGSVM [4] etc. The second one is the so-called one-off method [1],[5],[6]. It constructs k binary classifiers, where the m-th binary classifier separates the training vectors of class m from all the other vectors. Different from the above one-against-all approach, these k binary classifiers are obtained simultaneously by solving a global optimization problem at once. These two types of methods are complex and time-consuming. Both of their problem sizes are $O(k \times l)$ [1], where k is the number of classes, l is the number of samples. Thus, the computational time is increased linearly with the number of class k. On the other hand, the combination errors will be important as k becomes large, which will result in the quick decline of classifying accuracy. To reduce the computational time and combination errors, the third type of methods are proposed, which construct directly some multi classifier without the combination of binary classifiers. The large margin principle method [7] is just such an example, with problem size being only $O(2 \times l)$ and independent of the class number k. Because of this outstanding characteristic, the direct method has attracted much attention in ranking multi classification.

In practical classification problem, it often occurs that the importance of margins between classes is different. For example, the margin between class 1 and class 2 is more important than that between class 2 and class 3. The large margin principle method proposed in [7] has not yet considered the unequal margin importance. In this paper, we will extend the strategy of maximizing the sum of margin (MSM) in [7] to maximizing the sum of weighted margin (MSWM), so that we can deal with the ranking multi classification problem with unequal margin importance.

A linear SVM (LSVM) was proposed in [5] to solve the binary classification. As a linear model, it is relatively simple to be solved, and it is still of positive generalization ability. Is there any similar linear model for the multi classification problem? To answer this question, we introduce the concept of 'Algebraic Margin', and then propose a strategy of maximizing the sum of weighted algebraic margin (MSWAM). Based on MSWAM, a linear programming model for multi classification problem is obtained, which is similar to that for the binary classification.

This paper is organized as follows. In Section 2, we briefly introduce the direct approach for multi classifying problem in [7]. Section 3 is the main part of this paper. We first present MSWM, and deduce out its optimization model as well as its dual model. A numerical example is given to claim the effect of the models. Then we obtain a linear programming model by extending the maximum margin principle to the maximum algebraic margin principle, and point out the possibility that the solution of the problem based on MSWM may be gained by solving this linear programming problem. Finally, some conclusions and discussions are given in Section 4.

II. A DIRECT APPROACH FOR MULTI CLASSIFYING PROBLEM

We consider ranking multi classifying problem. Suppose $\{x_i^j\}$ is the set of training examples, where $j = 1,2,\cdots,k$ denotes the class number, and $i = 1,2,\cdots,n_j$ is the index within each class. Set $l = \sum_j n_j$, which denotes the total number of training examples. According to the maximum margin principle of SVM, we look for $(k-1)$ parallel separating hyper-planes

$$(w,b_1),\cdots,\ (w,b_{k-1}),\ b_1 \le b_2 \le \cdots \le b_{k-1}$$

such that the data are separated to k classes by the decision rule

$$f(x) = \min_{r\in\{1,2\cdots,k-1\}} \{r : w\cdot x - b_r < 0\}.$$

In other words, if input vector x satisfies $b_{r-1} < w\cdot x < b_r$, then x is assigned to class r, where $b_0 = -\infty$ and $b_k = +\infty$. The total number of margins is $k-1$ for this problem. There are two strategies to realize maximum margin principle for multi classifying problem: one is the strategy of maximum minimal margin, while the other is the strategy of maximizing the sum of margins (MSM). The objective of the first strategy is

$$\max_{w,b} 2/\|w\|,$$

where $2/\|w\|$ is the margin between the closest pair of classes. In MSM, the objective is

$$\max_{w,a_j,b_j} \sum_{j=1}^{k-1}(b_j - a_j)/\|w\|,$$

where

$$(w,a_1)\ ,(w,b_1),\cdots,(w,a_{k-1}),\ (w,b_{k-1}),$$
$$a_1 \le b_1 \le a_2 \le b_2 \le \cdots a_{k-1} \le b_{k-1}$$

are $2(k-1)$ parallel hyper-planes such that each class is sandwiched between two hyper-planes,

$$\sum_{j=1}^{k-1}(b_j - a_j)/\|w\|$$

denotes the sum of $k-1$ margins (See Fig.1.). If we normalize w by $\|w\| = 1$, then the objective of MSM is equivalent to

$$\min_{w,a_j,b_j} \sum_{j=1}^{k-1}(a_j - b_j).$$

In this paper, we only consider this extension of MSM.

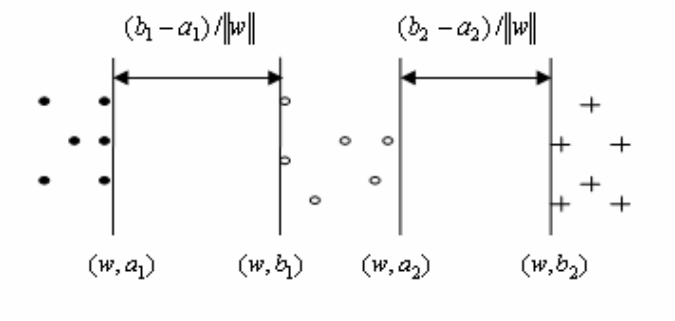

Figure 1. The strategy of maximizing the sum of margins

III. THE STRATEGY OF MAXIMIZING THE SUM OF WEIGHTED MARGINS

A. A Strategy of Maximizing the Sum of Weighted Margins

Considering the different demands of margins between classes, we extend MSM to maximizing the sum of weighted margins (MSWM). Suppose that the weights are ρ_j, $j = 1,2,\cdots k-1$, we can gain its programming problem ("soft margin") as follows:

$$\min_{w,a_j,b_j} \sum_{j=1}^{k-1}\rho_j(a_j - b_j) + C\sum_{j=1}^{k-1}(\sum_{i=1}^{n_j}\varepsilon_i^{j\#} + \sum_{i=1}^{n_{j+1}}\varepsilon_i^{j\&}) \quad (1)$$

s.t. $a_j \le b_j, j = 1,2,\cdots k-1$ (2)

$$b_j \le a_{j+1}, j = 1,2,\cdots k-2 \quad (3)$$

$$w\cdot x_i^j \le a_j + \varepsilon_i^{j\#},\ j = 1,\cdots,k-1,\ \ i = 1,\cdots,n_j \quad (4)$$

$$\varepsilon_i^{j\#} \ge 0, j = 1,\cdots,k-1,\ \ i = 1,\cdots,n_j \quad (5)$$

$$w\cdot x_i^{j+1} \ge b_j - \varepsilon_i^{j+1\&},\ j = 1,\cdots,k-1,\ \ i = 1,\cdots,n_{j+1} \quad (6)$$

$$\varepsilon_i^{j+1\&} \ge 0, j = 1,\cdots k-1,\ \ i = 1,\cdots n_{j+1} \quad (7)$$

$$w\cdot w \le 1\ , \quad (8)$$

where C denotes the penalty coefficient for misclassification, $\varepsilon_i^{j\#}$ and $\varepsilon_i^{j\&}$ are slack variables.

Remark 1. According to the original problem, condition (8) should be $w\cdot w = 1$. But here, it is natural to be extended to $w\cdot w \le 1$. (The reason is the same as [7], see [7] for details)

The model (1)-(8) is a convex programming problem, the total number of constraints is $(2k - 2 + 2l_1 + 2l_2)$, where

$$l_1 = \sum_{j=1}^{k-1} n_j\ ,\ l_2 = \sum_{j=2}^{k} n_j\ .$$

Theorem 1. The dual problem of the original model (1)-(8) is

$$\max_{\lambda}\ -\lambda H\lambda^T \quad (9)$$

s.t. $0 \le \lambda^{(1)} \le C^{(1)}, 0 \le \lambda^{(2)} \le C^{(2)}$ (10)

$$\vec{1}\cdot\lambda_1^{(1)} \ge \rho_1, \vec{1}\cdot\lambda_{k-1}^{(2)} \ge \rho_{k-1}, \vec{1}\cdot\lambda^{(1)} = \vec{1}\cdot\lambda^{(2)} \quad (11)$$

$$\sum_{i=1}^{j}\vec{1}\cdot\lambda_i^{(1)} \ge \sum_{i=1}^{j}\vec{1}\cdot\lambda_i^{(2)}, j = 1,2,\cdots,k-2 \quad (12)$$

$$\sum_{i=1}^{j+1}\vec{1}\cdot\lambda_i^{(1)} - \sum_{i=1}^{j}\vec{1}\cdot\lambda_i^{(2)} \ge \rho_{j+1}, j = 1,2,\cdots,k-3, \quad (13)$$

where

$$\lambda = [\lambda^{(1)},\lambda^{(2)}]_{1\times(l_1+l_2)}$$

$$H = \begin{bmatrix} (x_i^{(1)}\cdot x_j^{(1)}) & -(x_i^{(1)}\cdot x_j^{(2)}) \\ -(x_i^{(2)}\cdot x_j^{(1)}) & (x_i^{(2)}\cdot x_j^{(2)}) \end{bmatrix}_{(l_1+l_2)\times(l_1+l_2)},$$

$$\lambda^{(1)}=[\lambda_1^{(1)},\cdots,\lambda_{k-1}^{(1)}],\ \lambda_j^{(1)}=[\lambda_1^{j\#},\cdots,\lambda_{n_j}^{j\#}],$$
$$\lambda^{(2)}=[\lambda_1^{(2)},\cdots,\lambda_{k-1}^{(2)}],\ \lambda_j^{(2)}=[\lambda_1^{j+1\&},\cdots,\lambda_{n_{j+1}}^{j+1\&}],$$
$$x^{(1)}=[x_1^{(1)},\cdots,x_{k-1}^{(1)}],\ x_j^{(1)}=[x_1^{j},\cdots,x_{n_j}^{j}],$$
$$x^{(2)}=[x_1^{(2)},\cdots,x_{k-1}^{(2)}],\ x_j^{(2)}=[x_1^{j+1},\cdots,x_{n_{j+1}}^{j+1}],$$
$$C^{(1)}=[C,\cdots,C]_{1\times l_1},\ C^{(2)}=[C,\cdots,C]_{1\times l_2},$$

and $\vec{1}$ is the vector in which all elements are 1.

Proof. The Lagrangian function of problem (1)-(8) is

$$\begin{aligned}L(\cdot)=&\sum_{j=1}^{k-1}\rho_j(a_j-b_j)+C\sum_{j=1}^{k-1}(\sum_{i=1}^{n_j}\varepsilon_i^{j\#}+\sum_{i=1}^{n_{j+1}}\varepsilon_i^{j+1\&})\\&+\sum_{j=1}^{k-1}\psi_j(a_j-b_j)+\sum_{j=1}^{k-2}\eta_j(b_j-a_{j+1})\\&+\sum_{j=1}^{k-1}\sum_{i=1}^{n_j}\lambda_i^{j\#}(w\cdot x_i^j-a_j-\varepsilon_i^{j\#})\\&+\sum_{j=1}^{k-1}\sum_{i=1}^{n_{j+1}}\lambda_i^{j+1\&}(b_j-w\cdot x_i^{j+1}-\varepsilon_i^{j+1\&})\\&+\alpha(w\cdot w-1)\\&-\sum_{j=1}^{k-1}\sum_{i=1}^{n_j}\xi_i^{j\#}\varepsilon_i^{j\#}-\sum_{j=1}^{k-1}\sum_{i=1}^{n_{j+1}}\xi_i^{j+1\&}\varepsilon_i^{j+1\&}\end{aligned}\quad,\ (14)$$

where $\psi_j,\eta_j,\lambda_i^{j\#},\lambda_i^{j\&},\alpha,\xi_i^{j\#},\xi_i^{j\&}$ are nonnegative lagrangian multipliers. For convenience, we denote

$$\rho=[\rho_1,\rho_2,\cdots\rho_{k-1}],\quad \psi=[\psi_1,\psi_2,\cdots\psi_{k-1}],$$
$$a=[a_1,a_2,\cdots a_{k-1}],\quad \mathrm{b}=[\mathrm{b}_1,\mathrm{b}_2,\cdots\mathrm{b}_{k-1}],$$
$$\varepsilon^{(1)}=[\varepsilon_1^{(1)},\cdots\varepsilon_{k-1}^{(1)}],\quad \varepsilon_j^{(1)}=[\varepsilon_1^{j\#},\cdots,\varepsilon_{n_j}^{j\#}],$$
$$\varepsilon^{(2)}=[\varepsilon_1^{(2)},\cdots,\varepsilon_{k-1}^{(2)}],\quad \varepsilon_j^{(2)}=[\varepsilon_1^{j+1\&},\cdots,\varepsilon_{n_{j+1}}^{j+1\&}],$$
$$\xi^{(1)}=[\xi_1^{(1)},\cdots,\xi_{k-1}^{(1)}],\quad \xi_j^{(1)}=[\xi_1^{j\#},\cdots,\xi_{n_j}^{j\#}],$$
$$\xi^{(2)}=[\xi_1^{(2)},\cdots,\xi_{k-1}^{(2)}],\quad \xi_j^{(2)}=[\xi_1^{j+1\&},\cdots,\xi_{n_{j+1}}^{j+1\&}].$$

Set

$$\nabla_w L(\cdot)=0,\quad \nabla_a L(\cdot)=0,\quad \nabla_b L(\cdot)=0,$$
$$\nabla_{\varepsilon^{(1)}} L(\cdot)=0,\ \nabla_{\varepsilon^{(2)}} L(\cdot)=0,$$

we can gain

$$\sum_{j=1}^{k-1}\sum_{i=1}^{n_j}\lambda_i^{j\#}x_i^j-\sum_{j=1}^{k-1}\sum_{i=1}^{n_{j+1}}\lambda_i^{j+1\&}x_i^{j+1}+2\alpha w=0 \qquad (15)$$

$$\rho+\psi-[0,\eta_1,\cdots,\eta_{k-2}]-[\sum_{i=1}^{n_1}\lambda_i^{1\#},\cdots,\sum_{i=1}^{n_{k-1}}\lambda_i^{k-1\#}]=0 \qquad (16)$$

$$-\rho-\psi+[\eta_1,\cdots,\eta_{k-2},0]+[\sum_{i=1}^{n_2}\lambda_i^{2\&},\cdots,\sum_{i=1}^{n_k}\lambda_i^{k\&}]=0 \qquad (17)$$

$$[C,\cdots,C]-[\lambda_1^{(1)},\cdots,\lambda_{k-1}^{(1)}]-[\xi_1^{(1)},\cdots,\xi_{k-1}^{(1)}]=0 \qquad (18)$$

$$[C,\cdots,C]-[\lambda_1^{(2)},\cdots,\lambda_{k-1}^{(2)}]-[\xi_1^{(2)},\cdots,\xi_{k-1}^{(2)}]=0\ . \qquad (19)$$

From (15), we have

$$w=-\frac{1}{2\alpha}(\sum_{j=1}^{k-1}\sum_{i=1}^{n_j}\lambda_i^{j\#}x_i^j-\sum_{j=1}^{k-1}\sum_{i=1}^{n_{j+1}}\lambda_i^{j+1\&}x_i^{j+1})\ , \qquad (20)$$

where $\alpha>0$. In fact, if $\alpha=0$, then $\mathrm{w}\cdot\mathrm{w}\le 1$ is an unconstraint, which deduces that the solution of the original problem can not be achieved in its verge, and it will be derived to a contradiction with the result of [7]. Substitute *w* by (20) in the Lagrangian function, and notice formulas (16), (17), (18), (19), we have

$$L(\alpha,\lambda)=-\frac{1}{4\alpha}\lambda H\lambda^T-\alpha\ .$$

Set

$$\frac{\partial L(\alpha,\lambda)}{\partial\alpha}=-1+\frac{1}{4\alpha^2}\lambda H\lambda^T=0,$$

we can gain

$$\alpha=\frac{\sqrt{\lambda H\lambda^T}}{2}.$$

By substituting it to $\mathrm{L}(\alpha,\lambda)$, we have

$$L(\alpha,\lambda)=-\sqrt{\lambda\mathrm{H}\lambda^{\mathrm{T}}}\ .$$

Maximizing $-\sqrt{\lambda\mathrm{H}\lambda^{\mathrm{T}}}$ is equivalent to maximize $-\lambda\mathrm{H}\lambda^{\mathrm{T}}$, so we have (9). From (18) and (19), we can gain (10). ψ_j and η_j can be solved from (16) and (17). Then, by setting $\psi_j\ge 0,\eta_j\ge 0$, we can derive out (11),(12) and (13). Thus we finish our proof. □

It is similar to the elementary SVM that the normal vector *w* of the parallel hyper-planes is a combination of the support vectors (it can be easily seen from formula (20)) and all the thresholds a_j,b_j can also be gained by the support vectors. When there isn't any group of parallel hyper-planes to correctly separate the data in the input space, we can adopt a non-linear map, mapping the input space to a highly dimensional Hilbert space. Similar to the case of SVM, we don't need the details about the definition of this non-linear map, and what we need is just some kernel function $K(x_i,x_j)$.

Compared with the dual model (18)-(21) in [7], Theorem 1 increases the constraint conditions (12) and (13). Conditions (12) and (13) are necessary, and the nonnegative demand for the Lagrangian multipliers ψ_j and η_j will not be guaranteed on their absence.

B. A Numerical Example

We generate a dataset randomly in R^2, as shown in Fig.2, in which there are 3 classes of points. Suppose that the importance of margins between classes is different, for example, the importance of margin between Class '+' and Class 'o' is larger than that between Class 'o' and Class '*'. We solve this problem by using the above models. The results are listed in Tab.1 as well as in Fig. 2, where we denote M12 as the margin between Class '+' and Class 'o', M23 as the

Table.1 The margins for different weights

Weights	M12	M23	SWM
P1=0.5,p2=0.5	2.7650	1.3746	2.0698
P1=0.9,p2=0.1	2.7717	1.3246	2.6269

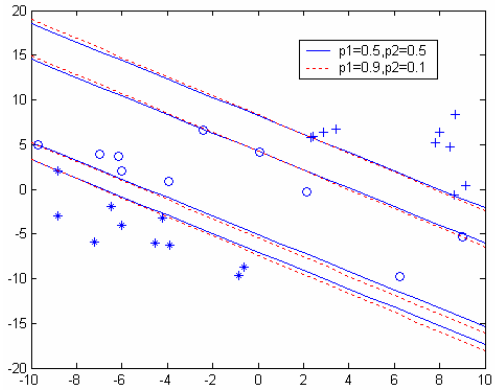

Figure 2. The results for different weights

margin between Class 'o' and Class '*', SWM as the sum of weight margins, p1 as the weight for M12, and p2 as the weight for M23.

From Tab.1 and Fig.2, it is clear that the bigger weight results in the bigger margin, and vice versa. For example, M12=2.7717 when p2=0.9, while M12=2.7650 when p2=0.5. These denote that the model based on MSWM can solve effectively the multi classification problems with unequal margin importance.

C. The Strategy of Maximizing the Sum of Weighted Algebraic Margins

Strictly speaking, the above margin $|b-a|/\|w\|$ stands for the geometric margin. Its optimization is a non-linear programming problem. When the problem size becomes large, its computational time is huge. Can we find a method that makes the computation easier (For example, much like a linear programming problem), and also has certain generalization ability? Here we propose a maximum algebraic margin principle. Under this principle, the mathematical model will be a linear programming problem, and with requested generalization ability.

Definition 1. Suppose that $(w,a),(w,b)$ are two arbitrary parallel hyper-planes and $x \in R^n$. The algebraic margin between point x and the hyper-plane (w,a) is $|w\cdot x-a|$, and the algebraic margin between the hyper-planes (w,a) and (w,b) is $|b-a|$.

For well distinguishing the two types of margins, we should generally call the margins used in the previous sections as the geometric margin. But for the simplicity, in this paper we say only 'margin' without the prefix 'geometric'. By its definition, the algebraic margin could be considered as some extension of the geometric margin (It plays the role of a geometric margin when $\|w\|$ =1). Although the algebraic margin has not the intuitive geometric significance, it possesses of certain generalization or anti-disturbance (include the disturbances of the input data and the separating hyper-plane) ability. By definition, we can see that the bigger the algebraic margin is, the better the generalization ability will be.

By extending the geometric margin to the algebraic margin, we can construct a model for the ranking multi classifying problem based on maximizing the sum of weighed algebraic margin (MSWAM) as follows:

$$\min_{w,a_j,b_j} \sum_{j=1}^{k-1}\rho_j(a_j-b_j)+C\sum_{j=1}^{k-1}(\sum_{i=1}^{n_j}\varepsilon_i^{j\#}+\sum_{i=1}^{n_{j+1}}\varepsilon_i^{j+1\&}) \tag{21}$$

$$\text{s.t.}\quad a_j \le b_j, j=1,2,\cdots k-1 \tag{22}$$

$$b_j \le a_{j+1}, j=1,2,\cdots k-2 \tag{23}$$

$$w\cdot x_i^j \le a_j+\varepsilon_i^{j\#},\ j=1,\cdots k-1,\ \ i=1,\cdots n_j \tag{24}$$

$$\varepsilon_i^{j\#}\ge 0, j=1,\cdots k-1,\ \ i=1,\cdots n_j \tag{25}$$

$$w\cdot x_i^{j+1} \ge b_j-\varepsilon_i^{j+1\&},\ j=1,\cdots k-1,\ \ i=1,\cdots n_{j+1} \tag{26}$$

$$\varepsilon_i^{j+1\&}\ge 0, j=1,\cdots k-1,\ \ i=1,\cdots n_{j+1} \tag{27}$$

$$w\ge -\vec{1}, w\le \vec{1}, \tag{28}$$

where $w\ge -\vec{1}$ and $w\le \vec{1}$ mean every element of w is not less than –1 and is not bigger than 1, which stands for the infinity-norm of w is not bigger than 1.

The above model is a linear programming model, which can be solved by many efficient methods. Considering that its feasible region includes the feasible region of problem (1)-(8), it might give an approach to solve the problem (1)-(8) by solving linear programming problems.

Ⅳ. CONCLUSION

In this paper, MSM is extended to MSWM. Using MSWM, a mathematical model and its dual model are derived out to deal with the ranking multi classification problem where the importance of margins between classes is different. The effect of the model is checked by a numerical example. Furthermore, by introducing the concept of algebraic margin, MSWM is extended to MSWAM. Based on MSWAM, the mathematical model not only has positive generalization ability, but is also a simple linear programming model. Hence, MSWAM has a practical meaning, and may be considered as a substitute of MSWM when the model based on MSWM suffers certain troubles.

The condition that a group of parallel hyper-planes separate the sample data corresponds to a stricter request on kernel function. How to select an adapted kernel function or achieve a relaxation for a group of parallel separating hyper-planes can be considered as a further work. In addition, the efficient algorithm for model (9)—(13), and the possible solving approach for model (1)—(8) by linear programming model (21)—(28), all are significative works in the future.

ACKNOWLEDGMENT

The authors thank Dr. W.Y Lan for many helpful suggestions. This work was supported by the "211" project, "Electronic Information and Technique", No: 0630-E11090, and the second phase of the "985" project, "Information Technique Platform", Xiamen University, P.R. China.

REFERENCES

[1] C.W. Hsu and C.J. Lin, "A comparison of methods for multiclass support vector machines", *IEEE Transactions on Neural Networks*, Vol.13,No.2,March 2002,pp.415-425.

[2] L. Bottou, C. Cortes, J. Denker, H. Drucker, I. Guyon, L. Jackel, Y. LeCun, U. Muller, E. Sackinger, P. Simard, and V. Vapnik, "Comparison of classifier methods: A case study in handwriting digit recognition," in *Proc. Int. Conf. Pattern Recognition*. , 1994, pp. 77–87.

[3] S. Knerr, L. Personnaz and G. Dreyfus, "Single-layer learning revisited: A stepwise procedure for building and training a neural network", in *Neurocomputing: Algorithms, Architectures and Applications* (J.Fogelman Ed.), New York: Springer-Verlag,1990.

[4] C.P. John, C. Nello and S.T. John, "Large margin DAGs for multiclass classification", *Advance in Neural Information Processing Systems*, Cambridge, MA:MIT Press,2000, Vol.12, pp.547-553.

[5] V.N. Vapnik, *Statistical Learning Theory*, New York: Wiley, 1998.

[6] J.Weston and C.Watkins, "Support vector machines for multi-class pattern recognition", *Proceedings of the 7th European Symposium on Artificial Neural Networks (ESANN-99)* (M.Verleysen Ed.), Bruges, Belgium, 1999, pp.219-224.

[7] A. Shashua and A. Levin, "Ranking with large margin principle:Two approaches", in *Advances in Neural Information Processing Systems*, MIT Press, 2003,Vol.15, pp:937-944.