

A Method of Optimizing Kernel Parameter of Sphere Structured Support Vector Machine

Kang Shouqiang^{1*}, Wang Yujing^{1,2}, Yang Guangxue¹, V.I. MIKULOVICH³

¹School of Electrical and Electronic Engineering, Harbin University of Science and Technology, Harbin, China

²School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin, China

³Belarusian State University, Minsk, Belarus

Abstract—Sphere structured support vector machine is a multi-classification algorithm. The algorithm separately constructs sphere for each class sample data, so the complex degree of the quadratic programming is reduced and it is easier to extend new samples. But, kernel parameter selection of sphere structured support vector machine needs to predetermine the parameter search interval. For eliminating human experience factors, a kernel parameter optimization method based on sphere center distance is proposed. Experimental results show that the proposed method can greatly shorten the training time in the case of the average classification accuracy of the classifier is not reduced.

Index Terms—support vector machine, sphere structured support vector machine, kernel method, pattern classification

I. INTRODUCTION

Support vector machine (SVM) is a new machine learning method based on statistical learning theory. The method shows unique advantages in solving small-sample problem [1][2]. Currently, many methods based on SVM have been presented, such as new fuzzy SVM [3]. Based on the proposed new fuzzy membership function, the method solves the sensitivity of the noise and outlier problems. Ref. [4] proposed a new example dependent costs SVM method, by using the method, more sensitive hyperplane can be obtained. Tax et al. [5] proposed the support vector domain description (SVDD) classification method for single-class data. For multi-classification problem, Ref. [6] proposed a sphere structured support vector machine multi-classification method. The method uses the hyper-sphere to define the same class data and the data space is composed of some hyper-spheres. This multi-classification method separately constructs hyper-sphere for each class sample data, so the complex degree of the quadratic programming is reduced, and when adding new class sample data, only need to construct the corresponding hyper-sphere, so it easier to extend new samples. SVM has been widely used in bearing fault diagnosis [7], clinical medicine [8], image classification

[9], and other research fields.

The parameters selection of SVM has developed some technologies: grid algorithm, cross validation method, trial and error procedures, evolutionary algorithm and gradient descent methods, and generalization error estimation [10][11]. But, all of these parameter selection methods need to predetermine the parameter search interval. The range size needs to be determined by human experience. Many numerical experiments and the experience show that when selecting Gaussian kernel function, the width parameter is the key factor for SVM model selection. The sphere structured support vector machine also has the same problem.

Aiming to the characteristics of the sphere structured support vector machine, the formula of the sphere center distance was derived. The method regards the sphere center distance as the separation index to determine the optimal selection range of kernel parameter of sphere structured multi-class support vector machine. The larger sphere center distance represents larger separation degree between two classes, then, the optimal kernel parameter range can be determined by the largest sphere center distance to reduce the training time. The propose method was tested on UCI datasets, the results shows that in the case of the same classification accuracy, optimized kernel parameter algorithm can reduce the training time of the sphere structured support vector machine. In the case of the same learning training consumption, if selecting more subtle kernel parameter, the classification accuracy could be improved possibly.

II. THE SPHERE STRUCTURED MULTI-CLASS SUPPORT VECTOR MACHINE

The principle of the sphere structured support vector machine is explained in detail in Ref. [6]. Mathematically, the m -class problem ($m > 2$) can be described as follows: Given n -dimensional element sets A^k , $k = 1, \dots, m$, each A^k contains l^k points x_i^k , $i = 1, \dots, l^k$, with all points belonging to the same class. For each set A^k , $k = 1, \dots, m$, we find a sphere (a^k, R_k) whose center is a^k and diameter is R_k . We minimize R_k so that the sphere (a^k, R_k) contains all or almost all sample points x_i^k , $i = 1, \dots, l^k$. Because this definition can be very sensitive to very outlying points, some points are allowed to fall outside the minimum hyper-sphere. The process of choosing the minimum

Supported by the Scientific and Technological Research Foundation of the Heilongjiang Provincial Department of Education 2011(No. 12511080)

*Corresponding author. Email:kangshouqiang@163.com

hyper-sphere can be transformed into an optimization problem:

$$\begin{cases} \min(R_k)^2 + C^k \sum_{i=1}^k \varepsilon_i^k \\ s. t. \begin{cases} \|x_i^k - a^k\|^2 \leq (R_k)^2 + \varepsilon_i^k, \\ \varepsilon_i^k \geq 0 \\ i = 1, 2, \dots, l^k \end{cases} \end{cases} \quad (1)$$

where ε_i^k is the introduced slack variable, and C^k is a certain constant that controls the punishment level for the misclassified samples, and achieves the trade-off between the size of the hyper-sphere and the misclassified samples.

Using the Lagrange multiplier method, above optimization problem can be transformed into the dual problem:

$$\begin{cases} \max L(\alpha_i^k) = \sum_i \alpha_i^k (x_i^k \cdot x_i^k) - \sum_{i,j} \alpha_i^k \alpha_j^k (x_i^k \cdot x_j^k) \\ s. t. \begin{cases} \sum_i \alpha_i^k = 1 \\ 0 \leq \alpha_i^k \leq C^k \\ i = 1, 2, \dots, l^k \end{cases} \end{cases} \quad (2)$$

where α_i^k is the Lagrange multiplier.

For m -class, we can obtain m hyper-spheres, the k^{th} hyper-sphere represents the k^{th} class. $D(z, a^k)$ is the square of the distance between the testing vector z and the k^{th} class sphere center a^k and is defined as

$$D(z, a^k) = \sum_{i,j} \alpha_i^k \alpha_j^k (x_i^k \cdot x_j^k) - 2 \sum_i \alpha_i^k (x_i^k \cdot z) + (z \cdot z). \quad (3)$$

The kernel function method similar to the ordinary SVM is applied in the sphere structured support vector machine. Using nonlinear mapping, the sample space is transformed into a higher dimensional space, so original boundary blur two type samples become can be separated. Equation (4) is obtained using the kernel function instead of the Euclidean dot product in (2).

$$\begin{cases} \max L(\alpha_i^k) = \sum_i \alpha_i^k K(x_i^k, x_i^k) - \sum_{i,j} \alpha_i^k \alpha_j^k K(x_i^k, x_j^k) \\ s. t. \begin{cases} \sum_i \alpha_i^k = 1 \\ 0 \leq \alpha_i^k \leq C^k \\ i = 1, 2, \dots, l^k \end{cases} \end{cases} \quad (4)$$

$D(z, a^k)$ is the square of the distance and is transformed into

$$D(z, a^k) = \sum_{i,j} \alpha_i^k \alpha_j^k K(x_i^k, x_j^k) - 2 \sum_i \alpha_i^k K(x_i^k, z) + K(z, z), \quad (5)$$

where $K(\cdot, \cdot)$ is the kernel function. Commonly used kernel functions are polynomial kernel function, radial basis kernel function and sigmoid function.

Based on (5), the classification rules of Ref. [6] are summarized as follows:

$$f(z) = \arg \min_{k=1}^m (D(z, a^k) - R_k^2). \quad (6)$$

Where $D(z, a^k)$ can be obtained by (5), $a^k = \sum_i \alpha_i^k x_i^k$,

$\alpha_i^k \neq 0$, $R_k^2 = D(z', a^k)$, z' is the support vector of the hyper-sphere surface.

III. THE KERNEL PARAMETER SELECTION OF THE SPHERE STRUCTURED CLASSIFIER AND THE ALGORITHM IMPLEMENTATION

Use the maximum sphere center distance method to optimize the kernel parameter of sphere structured multi-class SVM, that is, determine the optimal range of the kernel parameter. The sphere center distance is derived as follows:

Let it be supposed that the sphere center of the k_1^{th} class is a^{k_1} , the sphere center of the k_2^{th} class is a^{k_2} , d^2 is the square of two sphere center distance and is defined as

$$d^2 = D(a^{k_1}, a^{k_2}) = \|a^{k_1} - a^{k_2}\|^2 = (a^{k_1} \cdot a^{k_1}) - 2(a^{k_1} \cdot a^{k_2}) + (a^{k_2} \cdot a^{k_2}). \quad (7)$$

$a^k = \sum_i \alpha_i^k x_i^k$ is the sphere center of the k^{th} class hyper-sphere and is putted into (7), so

$$d^2 = \sum_i \sum_j \alpha_i^{k_1} \alpha_j^{k_1} (x_i^{k_1} \cdot x_j^{k_1}) - 2 \sum_i \sum_j \alpha_i^{k_1} \alpha_j^{k_2} (x_i^{k_1} \cdot x_j^{k_2}) + \sum_i \sum_j \alpha_i^{k_2} \alpha_j^{k_2} (x_i^{k_2} \cdot x_j^{k_2}), \quad (8)$$

where $i=1, 2, \dots, l_k; j=1, \dots, l_k$.

For nonlinear condition, the corresponding kernel function $K(\cdot, \cdot)$ is putted into (8), and the sphere center distance formula of two hyper-spheres is transformed into

$$d = \left[\sum_i \sum_j \alpha_i^{k_1} \alpha_j^{k_1} K(x_i^{k_1}, x_j^{k_1}) - 2 \sum_i \sum_j \alpha_i^{k_1} \alpha_j^{k_2} K(x_i^{k_1}, x_j^{k_2}) + \sum_i \sum_j \alpha_i^{k_2} \alpha_j^{k_2} K(x_i^{k_2}, x_j^{k_2}) \right]^{\frac{1}{2}}, \quad (9)$$

where $i=1, 2, \dots, l_k; j=1, \dots, l_k$.

The kernel function is used for mapping the sample space into a higher dimensional space, then, we can

obtain good kernel parameter which makes better separation effect between two type hyper-spheres. The relation between the sphere center distance value of the hyper-spheres and the kernel parameter is same to (9). So, the sphere center distance value of the hyper-spheres can represent the separation degree between two classes, and can be regarded as a class separability measurement in feature space. Based on the derived formula of sphere center distance, the optimal kernel parameter selection range can be determined by the maximum method of multi-class sphere center distance.

For multi-classification problem, the flow chart of the optimal kernel parameter selection range is shown as Fig. 1.

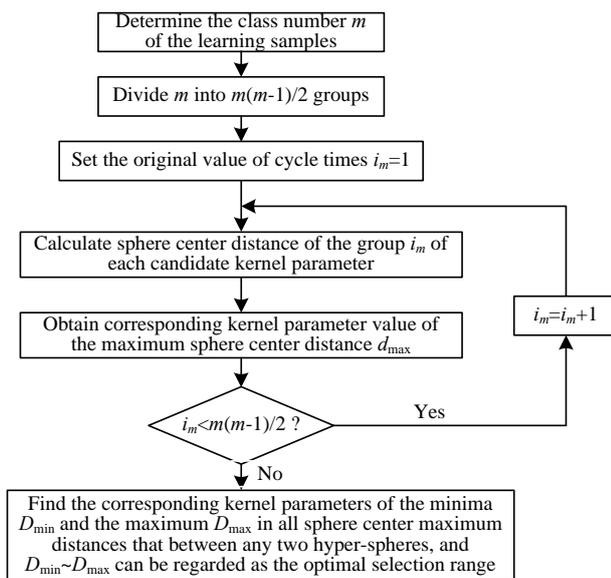


Figure 1. Flow chart of the optimal selection range of kernel parameter.

IV. THE EXPERIMENTS AND THE RESULTS ANALYSIS

All data were provided by UCI [12] data which covered dataset Iris, Ionosphere, Liver, New-thyroid, Sonar, Tae and Glass. The total class number of dataset samples is

two classes or three classes or six classes. Using cross-validation method, all experiments were tested by Matlab7.1 under Windows XP platform. When determining the hyper-sphere for sphere structured multi-class SVM, by the restriction condition in (4), So the value of C should be from 1// to 1. Kernel function was chosen as radial basis kernel function: $K(x_i, x_j) = \exp[-(x_i - x_j)^2 / 2s^2]$, where s was the kernel parameter. If s went to zero, then all sample points were support vector; if s went to infinity, then the discriminant function of radial basis kernel function SVM was a constant, that is, all sample points were classified as a class.

If the optimal value of s was sought by searching way, then the artificial selection range of s was very wide. Here s was restricted from 0.1 to 100 by the artificial experience, and the search step length was 1. The relations between classification parameters C, s and the classification precision were shown in Fig. 2(a), Fig. 3(a) and Fig. 4(a) for datasets Iris, New-thyroid, Glass. Through these figures, we could obtain the values of the parameters C and s under the highest average classification rate. The specific optimal s parameter value and the average classification accuracy were shown in Table 1. In order to validate the proposed kernel parameter s optimization method using hyper-sphere sphere center distance, the corresponding kernel parameter s_{min} and s_{max} for the minimum distance D_{min} and the distance interval D_{max} were obtained by calculating hyper-sphere sphere center distances of each class sample of each dataset, so the optimal selection range of kernel parameter s was $s_{min} \sim s_{max}$. The corresponding classification accuracy of every parameter value was shown in Fig. 2(b), Fig. 3(b) and Fig. 4(b). The specific values of s selection range, the step length and the average classification accuracy were shown in Table 1.

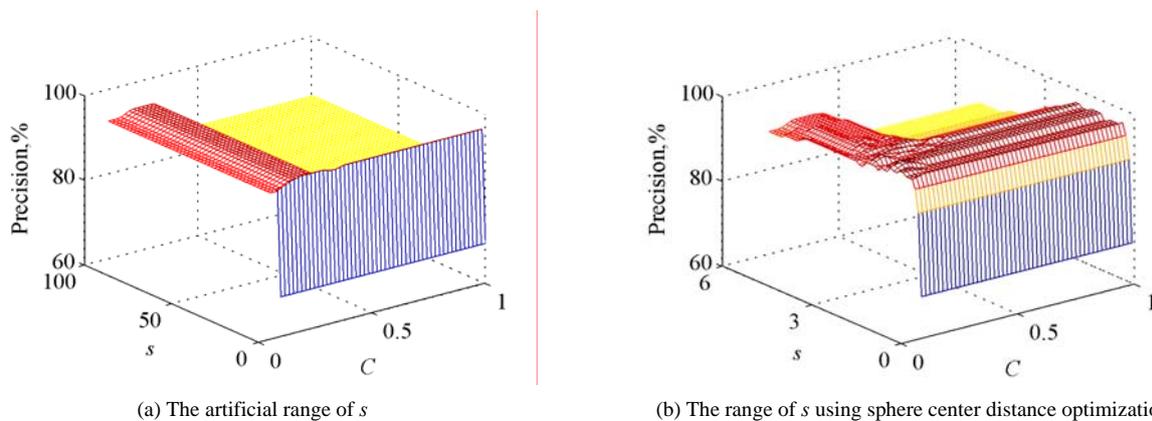
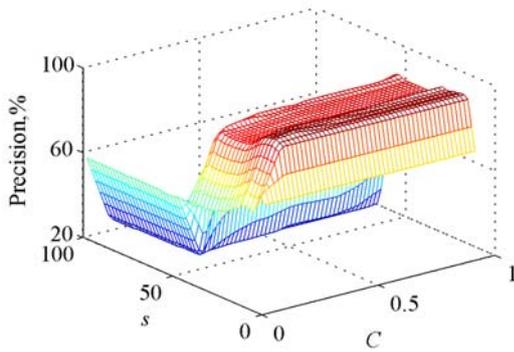
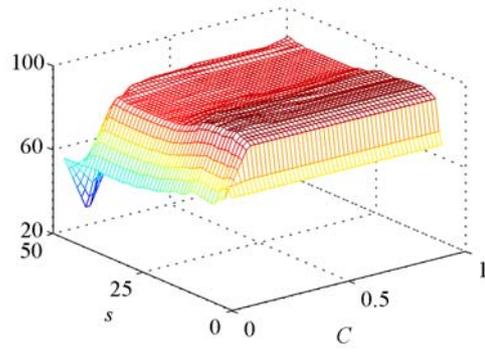


Figure 2. The relations between the parameters C, s and the classification accuracy (Iris dataset).

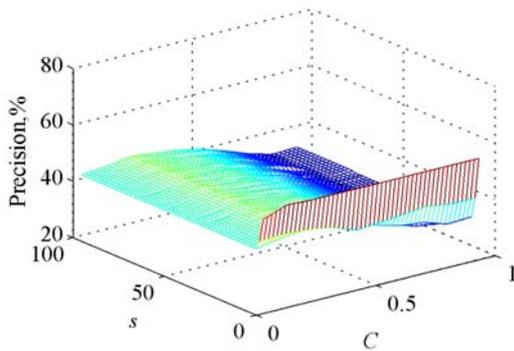


(a) The artificial range of s

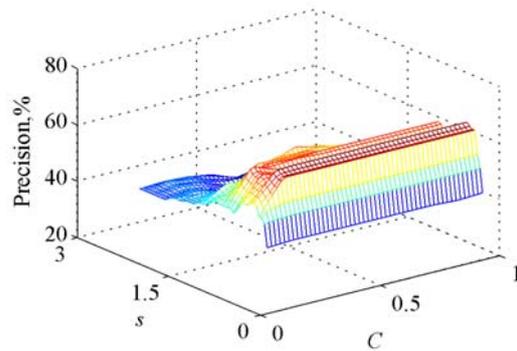


(b) The range of s using sphere center distance optimization

Figure 3. The relations between the parameters C , s and the classification accuracy (New-thyroid dataset)



(a) The artificial range of s



(b) The range of s using sphere center distance optimization

Figure 4. The relations between the parameters C , s and the classification accuracy (Glass dataset)

TABLE I.
THE CLASSIFICATION PARAMETER SELECTION AND THE AVERAGE CLASSIFICATION ACCURACY

Dataset	Total number of the data	The number of the class	The dimension of a datum	C	The artificial range of s (step length) (optimal s value)	The range of s using sphere center distance optimization $s_{min} \sim s_{max}$ (step length) (optimal s value)	The average classification accuracy using artificial s range /%	The average classification accuracy using optimized s range /%
Iris	150	3	4	0.15	[0.1-100] (1) (1.10)	[0.1-5.2] (0.052) (0.984)	95.33	96.67
Ionosphere	351	2	34	0.05	[0.1-100] (1) (2.10)	[0.1-5] (0.05) (1.96)	90.31	90.31
Sonar	208	2	60	0.015	[0.1-100] (1) (1.10)	[0.1-4.5] (0.045) (0.73)	75.48	76.92
New-thyroid	215	3	5	0.10	[0.1-100] (1) (8.10)	[1-48] (0.48) (8.20)	92.3	92.7
Liver	345	2	6	0.03	[0.1-100] (1) (12.10)	[0.1-56] (0.56) (11.86)	64.35	64.64
Tae	151	3	5	0.07	[0.1-100] (1) (0.10)	[0.1-8] (0.08) (0.42)	66.23	67.55
Glass	214	6	9	0.48	[0.1-100] (1) (1.10)	[0.01-2.1] (0.02) (0.27)	60.00	66.36

In Table 1, for different datasets (different total numbers of data, different numbers of class, different dimensions), the proposed hyper-sphere sphere center distance was regarded as the separation index, the obtained corresponding kernel parameters s_{min} and s_{max} of the minimum distances D_{min} and the maximum distances D_{max} were different, but the range of s became quite narrower than [0.1-100] which selected by the artificial experience. So, under the condition of the same search

step length, the training time of the classifier was reduced greatly, and the degree of time reduction depended on the reduction degree of optimized selection range. From another aspect to analyze, if the step lengths of two search ways was different, search times was almost the same, that is the same training time (such as Iris dataset, two search step lengths were 1 and 0.052 respectively), then the average classification accuracy under optimized s range was at least equal to the condition under

experiential selection s range, sometimes was higher a little, sometimes was higher greatly, and all of these were shown in Table 1.

V. CONCLUSION

The advantages and deficiencies are analyzed in the paper for sphere structured SVM. Aiming to the experience factors existing in kernel function parameter selection, an optimized kernel parameter selection method for sphere structured SVM is proposed. Obtain the maximum distances between every two hyper-sphere sphere centers of each class sample, and the optimal selection range of kernel parameter can be determined. Several datasets of UCI were test and the experimental results show that, by using the proposed kernel parameter optimization method, the narrow and effective search interval of the kernel parameter can be determined, and the experience factors are reduced. Therefore, under the condition of the same step length, the average accuracy is not reduced and the training time is reduced greatly.

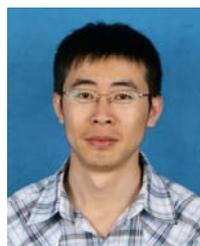
ACKNOWLEDGMENT

We would like to thank Ph.D. Kenneth A. Loparo (Case Western Reserve University) for his experimental data provided.

REFERENCES

[1] Vladimir N, "Statistical learning theory", *Wiley-Interscience*, New York, 1998.
 [2] Vladimir N, "The nature of statistical learning theory", *Springer-Verlag*, New York, 1999.
 [3] Yan Wei, and Xiao Wu, "A new fuzzy SVM based on the posterior probability weighting membership", *Journal of computers*, vol. 7, no. 6, pp. 1385–1392, 2012.
 [4] Xin Jin, Yujian Li, Yihua Zhou, and Zhi Cai, "Applying average density to example dependent costs SVM based on data distribution", *Journal of computers*, vol. 8, no. 1, pp. 91–96, 2013.
 [5] Tax D M J, and Duin R P W, "Support vector domain description", *Pattern Recognition Letters*, vol. 20, no. 11–13, pp. 1191–1199, 1999.
 [6] Zhu Meilin, Wang Yue, Chen Shifu, and Liu Xiangdong, "Sphere-structured support vector machines for multi-class pattern recognition", *Lecture Notes in Computer Science*, vol. 2639, pp. 589–593, 2003.

[7] Xuejun Li, Ke Wang, and Lingli Jiang, "The application of AE signal in early cracked rotor fault diagnosis with PWVD and SVM", *Journal of software*, vol. 6, no. 10, pp. 1969–1976, 2011.
 [8] Shanxiao Yang, and Guangying Yang, "Emotion recognition of EMG based on improved L-M BP neural network and SVM", *Journal of software*, vol. 6, no. 8, pp. 1529–1536, 2011.
 [9] Xijun Zhu, Dazhuan Liu, Qiulin Zhang, Zhaoshan Zhou, and Wenhua Liang, "Research of thenar palmprint classification based on gray level co-occurrence matrix and SVM", *Journal of computers*, vol. 6, no. 7, pp. 1535–1541, 2011.
 [10] Zhang Xiaoli, Chen Xuefeng, and He Zhengjia, "An ACO-based algorithm for parameter optimization of support vector machines", *Expert Systems with Applications*, vol. 37, no. 9, pp. 6618–6628, 2010.
 [11] Guangbin Wang, Yilin He, and Kuanfang He, "Multi-Layer kernel learning method faced on roller bearing fault diagnosis", *Journal of software*, vol. 7, no. 7, pp. 1531–1538, 2012.
 [12] UCI Machine Learning Repository [DB/OL]. (1987) [2013]. <http://www.ics.uci.edu/~mllearn/MLRepository.html>.



Kang Shouqiang got Ph. D. from Belarusian State University, Minsk, Belarus, in 2011. Now he is an associate professor and supervisor for M. Sc. students in Harbin University of Science and Technology, Harbin, China.

His main research interest is pattern classification and non-stationary signal processing technology.



Wang Yujing received M. Sc. from Harbin University of Science and Technology, Harbin, China, in 2007. She is currently a Ph. D. student in Harbin Institute of Technology, Harbin, China. And she is a lecturer in Harbin University of Science and Technology, Harbin, China.

Her main research interest is non-stationary signal processing and fault

diagnosis technology.