

Twin Support Vector Machines Based on the Mixed Kernel Function

Fulin Wu¹

¹School of Computer Science and Technology, China University of Mining and Technology, Xuzhou, China
Email: 562185156@qq.com

Shifei Ding^{1,2}

¹School of Computer Science and Technology, China University of Mining and Technology, Xuzhou, China
²Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100190 China
Email: dingsf@cumt.edu.cn

Abstract—The efficiency and performance of the Twin Support Vector Machines (TWSVM) are better than the traditional support vector machines when it deals with the problems. However, it also has the problem of selecting kernel functions. Generally, TWSVM selects the Gaussian radial basis kernel function. Although it has a strong learning ability, its generalization ability is relatively weak. In a certain extent, this will limit the performance of TWSVM. In order to solve the problem of selecting kernel functions in TWSVM, we propose the twin support vector machines based on the mixed kernel function (MK-TWSVM) in this paper. To make full use of the learning ability of local kernel functions and the excellent generalization ability of global kernel functions, MK-TWSVM selects a global kernel function and a local kernel function to construct a mixed kernel function which has the better performance. The experimental results indicate that the mixed kernel function makes TWSVM have the good learning ability and generalization ability. So it improves the performance of TWSVM.

Index Terms—mixed kernel function; TWSVM; kernel function

I. INTRODUCTION

Support Vector Machine (SVM) was proposed by Vapnik [1-3] et al firstly. It is based on the VC dimension theory and the principle of structural risk minimization in the statistical learning theory [4-5]. It has been applied in many fields [6-11] and there have been many improvements [12]. In 2001, Fung and Mangasarian [13] proposed the Proximal Support Vector Machines (PSVM). PSVM uses the equality constraints instead of the inequality constraints in the traditional SVM to make the calculation of PSVM simple. But for the points near the separating hyperplane, the classification accuracy is

insufficient. In 2006, based on the study of PSVM, Proximal SVM based on Generalized Eigenvalues (GEPSVM) was proposed by Mangasarian [14] et al. GEPSVM cancels the constraint that the two hyperplanes must be parallel in PSVM. GEPSVM makes each type of sample points as close as possible to its hyperplane and as far away as possible from the other sample points. Further, the solution of the problem is converted to the solution of the smallest eigenvalue of the two generalized eigenvalue problems to obtain the global extremum [15]. Thereafter, in 2007, based on the PSVM and GEPSVM, Jayadeva [16] et al proposed Twin Support Vector Machines (TWSVM). TWSVM solves a hyperplane for each type of sample points and makes each type of sample points as close as possible to its hyperplane and as far away as possible from another type of sample points' hyperplane. The two hyperplanes in TWSVM have no constraint on the parallel condition. The binary classification problem is converted to two smaller quadratic programming problems by TWSVM.

Because TWSVM has the solid theoretical foundation and the superiority of solving problems, many scholars have contributed to the study of TWSVM since TWSVM was proposed [17-18]. There have been many achievements in the efforts of research workers. For example, Jing Chen [19] proposed Weighted Least Squares TWSVM (WLSTWSVM), Qi Zhiqian [20] proposed a new type of Robust Twin Support Vector Machine for pattern classification, in 2009, Xinsheng Zhang [21] et al applied TWSVM to the detection of MCs.

However, all of these improved algorithms have the problem of selecting kernel functions consistently. The selection of the kernel function will affect the performance of the algorithm directly. Most algorithms only select a global kernel function or a local kernel function. However, both the global kernel function and the local kernel function have certain deficiencies. The global kernel function has a good generalization ability, but its learning ability is relatively weak. The local kernel

Manuscript received August 1, 2013; revised September 2, 2013; accepted September 28, 2013.

This work is supported by the National Natural Science Foundation of China (No.61379101), and the National Key Basic Research Program of China (No.2013CB329502).

Corresponding author, Shifei Ding, Email:dingsf@cumt.edu.cn.

function has a good learning ability, but its generalization ability is relatively weak. This will affect the performance of the algorithm. This paper proposes Twin Support Vector Machines based on Mixture Kernel Function (MK-TWSVM) to further improve the performance of TWSVM. MK-TWSVM uses a global kernel function and a local kernel function to construct a new kernel function. This mixed kernel function takes the learning ability and generalization ability into full account to find a best balance point among them. Since MK-TWSVM makes use of the learning ability of the local kernel function and the generalization ability of the global kernel function, it improves the performance of TWSVM.

The rest of this paper is organized as follows: Section II briefly describes the mathematical model of TWSVM and analyzes the learning ability and generalization ability of several commonly used kernel functions in detail. Section III constructs the mixed kernel function and describes MK-TWSVM. Section IV analyzes the experimental results in detail. Finally, we summarize and conclude the paper.

II. TWSVM

In 2007, Twin Support Vector Machines (TWSVM) was proposed by Jayadeva[16] et al. The solution of binary classification problem is converted to the solution of two smaller quadratic programming problems by TWSVM [14]. And then it gets two non-parallel hyperplanes. It makes each type of sample points as close as possible to its hyperplane and as far away as possible from another type of sample points' hyperplane. We use A and B to represent the two hyperplanes. If a sample point is closer to A, it belongs to the category which A represents. If a sample point is closer to B, it belongs to the category which B represents. Shown in Figure 1, the two lines represent the two classified hyperplanes, and the purple dots and green dots represent the training points of Category 1 and Category -1.

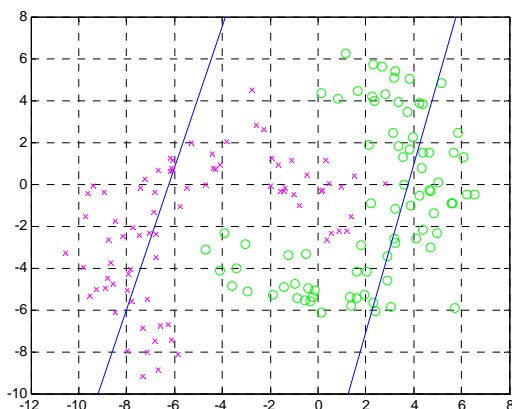


Figure.1 the basic idea of TWSVM

A. The mathematical model of TWSVM

We assume that there are l training samples in the space of R^n and they all have n attributes. m_1 samples of them are part of the positive class and m_2 samples of them are part of the negative class. We use the matrix of $A(m_1 \times n)$ and the matrix of $B(m_2 \times n)$ to represent them respectively. Finding two non-parallel hyperplanes in the space of R^n is the solving process of TWSVM:

$$x^T w_1 + b_1 = 0 \text{ and } x^T w_2 + b_2 = 0 \quad (1)$$

However, in the nonlinear separable case, we need to introduce the kernel function $K(x^T, C^T)$. At this time the two hyperplanes of TWSVM are as following:

$$K(x^T, C^T)w_1 + b_1 = 0 \text{ and } K(x^T, C^T)w_2 + b_2 = 0 \quad (2)$$

We construct the solution of the problems and it is as following:

$$\min \frac{1}{2} \|K(A, C^T)w_1 + e_1 b_1\|^2 + c_1 e_2^T \zeta \quad (3)$$

$$s.t - (K(B, C^T)w_1 + e_2 b_1) + \zeta \geq e_2, \zeta \geq 0, \quad (4)$$

$$\min \frac{1}{2} \|K(B, C^T)w_2 + e_2 b_2\|^2 + c_2 e_1^T \zeta \quad (5)$$

$$s.t - (K(A, C^T)w_2 + e_1 b_2) + \zeta \geq e_1, \zeta \geq 0, \quad (6)$$

In the above formula, $C^T = [A \ B]^T$, e_1 is the unit column vector which has the same number of rows with the kernel function of $K(A, C^T)$, e_2 is the unit column vector which has the same number of rows with the kernel function of $K(B, C^T)$. ζ is the slack vector, $A = [x_1^{(1)}, x_2^{(1)}, \dots, x_{m_1}^{(1)}]^T$, $B = [x_1^{(1)}, x_2^{(1)}, \dots, x_{m_1}^{(1)}]^T$, $x_j^{(i)}$ represents the j th sample in the i th class.

The distance between the test samples and the hyperplanes determines which category the test samples will be classified as. It means that if

$$K(x^T, C^T)w_r + b_r = \min_{i=1,2} \|K(x^T, C^T)w_i + b_i\|, \quad (7)$$

x belongs to the r th class and $r \in \{1,2\}$.

B. The kernel function

Data which are linearly inseparable in the low-dimensional space can be mapped into a high dimensional feature space by the kernel function to be linearly separable. And this avoids "the curse of dimensionality" when it computes in the high dimensional feature space.

Theorem 1 (Mercer) [22]:

When $g(x) \in L_2(R^N)$ and $k(x, x') \in L_2(R^N \times R^N)$, if $\iint k(x, x')g(x)g(x')dx dy \geq 0$ is right, we have $k(x, x') = (\Phi(x) \cdot \Phi(x'))$, it means that k is the inner product of the feature space.

According to the Mercer theorem, the following properties of kernel functions can be easily proved out. By the following properties, based on the common kernel functions, we can construct a new kernel function which we want. And this new kernel function performance can be used to improve our algorithm with the better performance.

Property 1 [23]: Let k_1 and k_2 are the kernel functions defined on $x \times x$ when $a \in R^+$. Then the following functions are kernels:

$$k(x, z) = k_1(x, z) + k_2(x, z) \tag{8}$$

$$k(x, z) = ak_1(x, z) \quad a > 0 \tag{9}$$

The most commonly used kernel functions are the linear kernel function, the polynomial kernel function, Gaussian radial basis kernel function and the sigmoid kernel function. Their expressions are as follows:

1. The linear kernel function:

$$K(x, x_i) = x \cdot x_i \tag{10}$$

2. The polynomial kernel function:

$$K(x, x_i) = (\gamma(x \cdot x_i) + r)^d, \gamma > 0 \tag{11}$$

3. The Gaussian radial basis kernel function:

$$K(x, x_i) = \exp\left(-\frac{\|x \cdot x_i\|^2}{2\sigma^2}\right) \tag{12}$$

4. The sigmoid kernel function:

$$K(x, x_i) = \tanh(\nu(x \cdot x_i) + c) \tag{13}$$

In TWSVM, once the kernel function and its parameters are determined, the model of TWSVM is determined. It evaluates the model of an algorithm with its learning ability and generalization ability. The kernel functions can usually be divided into two categories: global kernel functions and local kernel functions. The global kernel functions have a good generalization ability. Because it allows the data points which are very far away from each other to have an effect on the kernel function. But its learning ability is weak. The local kernel functions have a good learning ability, but its generalization ability is weak. That is because it only allows closely spaced data points to have an effect on the kernel function.

The next we will analyze the sigmoid kernel function which is one of the global kernel functions and Gaussian radial basis kernel function which is one of the local kernel functions.

The sigmoid kernel function is a common global kernel function and its expression is as following:

$$K(x, x_i) = \tanh(\nu(x \cdot x_i) + c) \tag{14}$$

Figure 2 is a graph at the test point of 0.1 when c in the

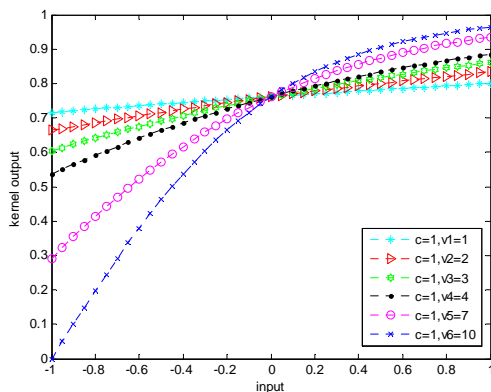


Figure.2 The graph of the sigmoid kernel function ($c = 1$) at the test point of 0.1

sigmoid kernel function is a fixed value and ν in the sigmoid kernel function has different values. From the Figure 2 we can see that the sigmoid kernel function has better results and generalization ability when ν has the value of 1 or 2.

Figure 3 is a graph at the test point of 0.1 when ν in the sigmoid kernel function has the value of 2 and c in the sigmoid kernel function has different values. From the Figure 3 we can see that the sigmoid kernel function has better results and generalization ability when $c \geq 4$, and the output values reach a steady state.

If we further analyze the Figure 3, we can know whether the data points near the test data point or not can

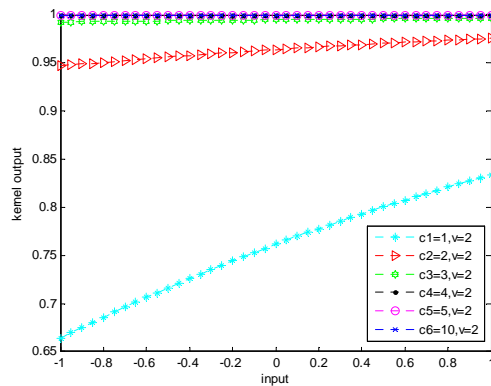


Figure.3 The graph of the sigmoid kernel function ($\nu = 2$) at the test point of 0.1

have an effect on the sigmoid kernel function, but at the test point its learning ability is relatively poor. After several experiments, we conclude that ν with the value of 2 is more appropriate.

Gaussian radial basis kernel function is a common local kernel function and its expression is as following:

$$K(x, x_i) = \exp\left(-\frac{\|x \cdot x_i\|^2}{2\sigma^2}\right) \tag{15}$$

Figure 4 is a graph of Gaussian radial basis kernel function at the test point of 0.1 when σ expressed in Figure 4 by p has different values. As it can be seen from Figure 4, Gaussian radial basis kernel function at the test point has a strong learning ability. But its generalization ability is relatively poor, because it only allows closely spaced data points to have an effect on the kernel function. We can see from the figure 4, the parameter value and learning ability are inversely proportional. That means the greater the parameter value is, the worse its learning ability will be. After a lot of experiments and practice, we know that the value of σ between 0.1 and 1 is better in general.

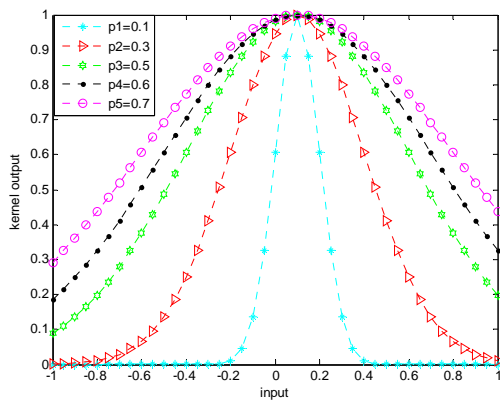


Figure.4 The graph of Gaussian radial basis kernel function at the test point of 0.1

III. USING THE MIXED KERNEL FUNCTION TO IMPROVE THE TWSVM

Although there are many improved algorithms of TWSVM including optimization of parameters in kernel functions, these algorithms don't fundamentally change the learning ability and generalization ability of the kernel functions to improve the performance of TWSVM. We take the improvement of the generalization and learning ability in the kernel function as a starting point, and then this paper proposes the twin support vector machines based on the mixed kernel function. This algorithm makes full use of the generalization ability of global kernel functions and the learning ability of local kernel functions to improve the performance of TWSVM.

A. The Mixed Kernel Function

This paper selects the sigmoid kernel function which is a common global kernel function and Gaussian radial basis kernel function which is a common local kernel function as the basic kernel functions to construct a mixed kernel function. Then it will be used to improve the performance of TWSVM. Based on this idea, we construct a function which is as following:

$$K(x, x_i) = aK_1(x, x_i) + bK_2(x, x_i), a > 0, b > 0 \quad (16)$$

In formula (16), $K_1(x, x_i)$ is the sigmoid kernel function and $K_2(x, x_i)$ is Gaussian radial basis kernel function.

The next, let us prove that this mixed function is an admissible kernel function.

Proof: According to the above-mentioned formula (9), $aK_1(x, x_i)$ with $a > 0$ is an admissible kernel function.

Similarly, $bK_2(x, x_i)$ with $b > 0$ is an admissible kernel function. Let $K_3(x, x_i) = aK_1(x, x_i)$ where $a > 0$ and

$K_4(x, x_i) = bK_2(x, x_i)$ where $b > 0$, then both

$K_3(x, x_i)$ and $K_4(x, x_i)$ are admissible kernel functions.

Let $K_5(x, x_i) = K_3(x, x_i) + K_4(x, x_i)$, then according to the above-mentioned formula (8), $K_5(x, x_i)$ is an admissible kernel function. $K_5(x, x_i)$ is just the $K(x, x_i)$. Therefore,

$K(x, x_i) = aK_1(x, x_i) + bK_2(x, x_i)$ with $a > 0$ and $b > 0$ is an admissible kernel function. QED.

a and b in the formula (16) represent the percentages of the sigmoid kernel function and Gaussian radial basis kernel function in the mixed kernel function. In order to ensure that the mixed kernel function does not change the reasonableness of the original mapping, generally, let $0 \leq a, b \leq 1$ and $a + b = 1$ [23]. According to this, the formula (16) can be converted to:

$$K(x, x_i) = \lambda K_1(x, x_i) + (1 - \lambda) K_2(x, x_i) \quad 0 \leq \lambda \leq 1 \quad (17)$$

In formula (17), $K_1(x, x_i)$ is the sigmoid kernel function and $K_2(x, x_i)$ is Gaussian radial basis kernel function. Therefore, the final mathematical expression of the mixed kernel function is as following:

$$K(x, x_i) = \lambda \tanh(\nu(x \cdot x_i) + c) + (1 - \lambda) \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) \quad 0 \leq \lambda \leq 1 \quad (18)$$

Let $s = 1 / 2\sigma^2$ then the formula (18) can be converted to:

$$K(x, x_i) = \lambda \tanh(\nu(x \cdot x_i) + c) + (1 - \lambda) \exp(-s \cdot \|x - x_i\|^2) \quad 0 \leq \lambda \leq 1 \quad (19)$$

Figure 5 is a graph of the mixed kernel function at the test point of 0.1 when $\nu = 2, c = 10$ and λ has different values. a in the Figure 5 represents the parameter λ in the mixed kernel function. It can be seen from the Figure 5 that the generalization ability and learning ability of the mixed kernel function will be different when λ has different values which means that the percentage of the sigmoid kernel function and Gaussian radial basis kernel function changes. The generalization ability of the mixed kernel function will be enhanced when the value of λ increases. For different data sets, λ will have different values to achieve the best results.

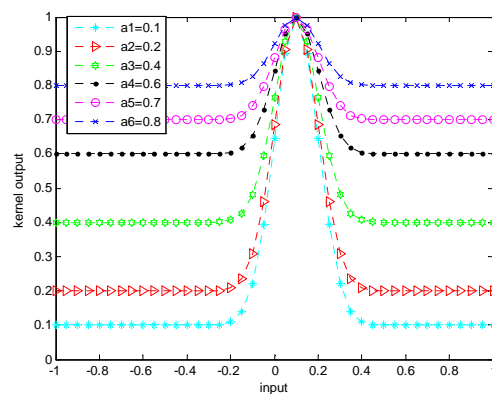


Figure.5 The graph of the mixed kernel function at the test point of 0.1

B. The Description of MK-TWSVM

MK-TWSVM is described as follows:

Step1 Import the data sets and divide these data sets into two randomly. One is 80% of the data, and the other is 20%.

Step 2 Set the parameters in the sigmoid kernel function and initialize the algorithm.

Step 3 Take the 80% of the data for training and determine the value of λ in the mixed kernel function, the value of σ in Gaussian radial basis kernel function and the value of c_1 、 c_2 in TWSVM by the grid searching method.

Step 4 Calculate the classification accuracy with the parameter values from **Step 3**.

Step 5 Determine whether it is the global optimum accuracy. If it is the global optimum accuracy, update the global optimum value and record this optimal parameter values. If it is not the global optimum accuracy, Jump to the **Step 6**.

Step 6 Determine whether it reaches the end condition of the grid cycle. If it does not, Jump to the **Step 3**. If it does, Jump to the **Step 7**.

Step 7 Bring the optimal parameters from the training into TWSVM. And then the final model of MK-TWSVM is determined.

Step 8 After the model of MK-TWSVM is determined, take the remaining 20% of the data for testing to get the test classification accuracy.

Step 9 Stop operations.

The flow chart of MK-TWSVM is shown in Figure 6. By this flow chart, we can intuitively understand the process of MK-TWSVM proposed by this paper. And the nine steps of the algorithm described above are clearly expressed in Figure 6. This will help us understand the algorithm.

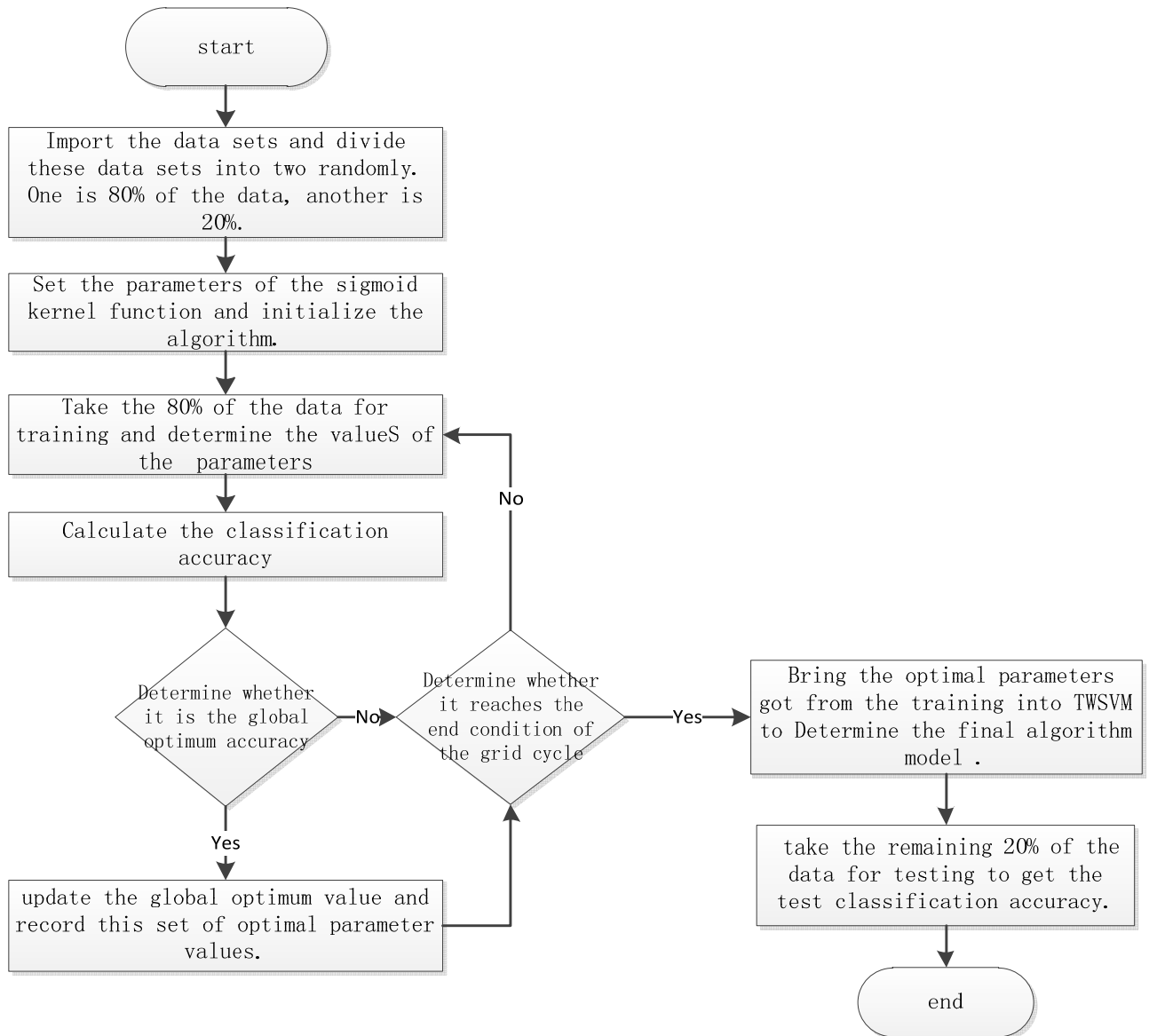


Figure.6 The flow chart of MK-TWSVM

IV. ANALYSIS OF THE EXPERIMENTAL RESULTS

This paper selects five common data sets in the UCI machine learning database to test and validate the algorithm proposed by us. The 80% of the data will be used for training and the remaining 20% of the data will be used for testing. Since we want to verify that the mixed kernel function proposed by this paper will improve the performance of TWSVM, we only do the nonlinear experiments. The five data sets are ionosphere data set, Sonar data set, votes data set, bupa data set and Hepatitis data set. Using the MATLAB environment, we do these experiments on a PC. In this algorithm, the value of ν in the sigmoid kernel function is 2 and the value of c in the sigmoid kernel function is 10. We will get the optimal values of λ in the mixed kernel function, s in Gaussian radial basis kernel function and c_1, c_2 in TWSVM through grid computing. For different data sets, their values are different.

The used data sets are sonar data set, ionosphere data set, Votes data set, bupa data set and Hepatitis data set in these experiments. Their characteristics are shown in Table I.

TABLE I.
THE DATA CHARACTERISTICS OF THE DATA SETS

Data sets	The number of samples	The number of attributes
Sonar	208	60
Ionosphere	351	34
Votes	435	16
bupa	345	7
Hepatitis	155	19

In the experiments, MK-TWSVM randomly selects 80% of the data to be used for training and then it gets the optimal parameters to determine the model. Then, the remaining 20% of the data is used for testing to get the corresponding classification accuracy. We compare the experimental results of MK-TWSVM with PSVM, GEPSVM and TWSVM. The Table II shows the results of the comparison.

TABLE II.
THE EXPERIMENTAL RESULTS

Data sets	MK-TWSVM	TWSVM	GEPSVM	PSVM
Sonar	93.02	89.64	85.97	82.79
Ionosphere	95.77	87.46	84.41	90.83
Votes	96.59	94.91	94.5	93.70
bupa	69.57	68.64	68.18	65.8
Hepatitis	82.35	81.39	79.28	78.57

In order to visually observe the experimental results, we plot the results in Figure 7. The ordinate represents the classification accuracy values. The abscissa of 1 represents the sonar data set. The abscissa of 2 represents the bupa data set. The abscissa of 3 represents the ionosphere data set. The abscissa of 4 represents the Hepatitis data set. The abscissa of 5 represents the Votes data set. The effect diagram is shown in Figure 7:

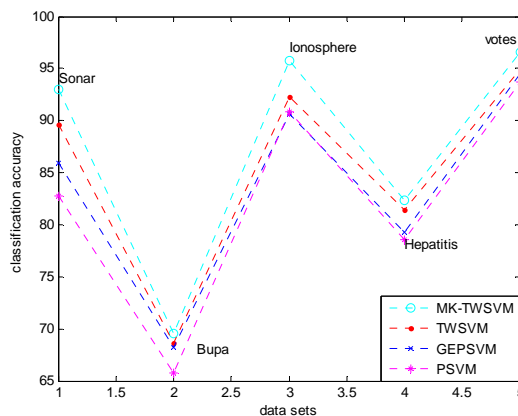


Figure.7 The effect diagram of the comparison

We can see from the test results that the classification accuracy of MK-TWSVM proposed by this paper has increased significantly, compared with the traditional classification algorithms. In Figure 7, we can see more intuitively that the classification accuracy curve of MK-TWSVM is obviously above the classification accuracy curves of TWSVM, GEPSVM and PSVM. It means that the classification accuracy of MK-TWSVM is better than theirs and has significantly improved. Why can it be able to achieve such a significant effect? It is because that MK-TWSVM has used the mixed kernel function. The mixed kernel function does not randomly select any kernel functions to be combined, but selects a global kernel function and a local kernel function to be combined. The global kernel function has a good generalization ability and the local kernel function has a good learning ability. MK-TWSVM adjusts the proportion of the global kernel function and the local kernel function through adjusting the value of λ to further make the learning ability and generalization ability achieve an optimum balance. We can see from this analysis that the biggest improvement of MK-TWSVM is the introduction of the mixed kernel function. And it makes full use of the advantages of the global and local kernel functions to improve the performance of TWSVM.

V. CONCLUSIONS

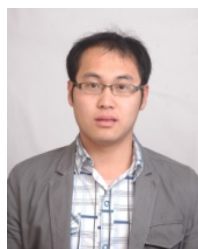
In recent years, classification algorithms have various improvements and TWSVM also has been developed rapidly. But TWSVM still has many defects, such as the problem of selecting the kernel function. For this problem, we propose the MK-TWSVM in this paper. It makes full use of the generalization ability of global kernel functions and the learning ability of local kernel functions and then it achieves an optimum balance between them to further determine the model of MK-TWSVM. This avoids the case that the traditional TWSVM only uses a single kernel function. In that case, there exists a problem. If it selects a global kernel function, its generalization ability is good, but the learning ability is relatively weak. If it selects a local kernel function, its learning ability is good, but the generalization ability is relatively weak. Therefore, MK-TWSVM selects a global kernel function and a local

kernel function to construct a mixed kernel function which has the better performance. And then MK-TWSVM introduces it into TWSVM. So MK-TWSVM can make full use of their respective advantages to improve the performance of TWSVM.

The experiments show that MK-TWSVM improves the classification accuracy of TWSVM. But MK-TWSVM also has a defect that the parameters are difficult to be determined. In addition to the parameters of TWSVM, there are parameters of the mixed kernel function. So MK-TWSVM relatively has more parameters and finds the optimal parameters more difficultly. Because of the more parameters, it will take more time. Therefore, we can start from this point in the following research work. If we can optimize the parameters, we will improve the performance of MK-TWSVM to further improve the classification efficiency and accuracy of TWSVM.

REFERENCES

- [1] CRISTIANINI N, TAYLOR J S. An introduction to support vector machines and other kernel-based learning methods. Translated by Li Guozheng, Wang Meng, Zeng Huajun. Beijing: Electronic Industry Press, 2004.
- [2] Ding Shifei, Qi Bingjuan, Tan Hongyan. An Overview on Theory and Algorithm of Support Vector Machines. Journal of University of Electronic Science and Technology of China, 2011, 40(1):2-10.
- [3] C. Cortes, V. Vapnik. Support-Vector Networks. Machine Learning, 1995, 20(2), 273-297.
- [4] Vapnik V N. The nature of statistical learning theory. Translated by Zhang Xuegong. Beijing: Tsinghua University Press, 2000.
- [5] Vapnik V.N. Statistical Learning Theory. Translated by Xu Jianhua, Zhang Xuegong. Beijing: Electronic Industry Press, 2004.
- [6] Masaki Murata, Tomohiro Mitsumori, Kouichi Doi. Analysis and Improved Recognition of Protein Names Using Transductive SVM. Journal of Computers, 2008, 3(1), 51-62.
- [7] Lianwei Zhang, Wei Wang, Yan Li, Xiaolin Liu, Meiping Shi, Hangen He. A Method for Surface Reconstruction Based on Support Vector Machine. Journal of Computers, 2009, 4(9), 806-812.
- [8] Qisong Chen, Xiaowei Chen, Yun Wu. Optimization Algorithm with Kernel PCA to Support Vector Machines for Time Series Prediction. Journal of Computers, 2010, 5(3), 380-387.
- [9] Wei Liu, Yuhua Yan, Rulin Wang. Application of Hilbert-Huang Transform and SVM to Coal Gangue Interface Detection. Journal of Computers, 2011, 6(6), 1262-1269.
- [10] Lei Zhang, Yanfei Dong. Research on Diagnosis of AC Engine Wear Fault Based on Support Vector Machine and Information Fusion. Journal of Computers, 2012, 7(9), 2292-2297.
- [11] Lei Ding, Fei Yu, Sheng Peng, Chen Xu. A Classification Algorithm for Network Traffic based on Improved Support Vector Machine. Journal of Computers, 2013, 8(4), 1090-1096.
- [12] J.A.K. SUYKENS, J. VANDEWALLE. Least Squares Support Vector Machine Classifiers. Neural Processing Letters, 1999, 9:293-300.
- [13] Fung G, Mangasarian O L. Proximal support vector machine classifiers. In: Proc 7th ACM SIGKDD Intl Conf on Knowledge Discovery and Data Mining, 2001: 77-86.
- [14] Mangasarian Olvi L, Wild Edward W. Multisurface proximal support vector machine classification via generalized eigenvalues. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(1): 69-74.
- [15] Li Kai, Lu Xiao Xia. Twin Support Vector Machine Algorithm with Fuzzy Weighting. Computer Engineering and Applications. 2011.
- [16] Jayadeva, Khemchandni Reshma, Suresh Chandra. Twin support vector machines for pattern classification. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(5):905-910.
- [17] Ding Shifei, Yu Junzhao, Qi Bingjuan. An Overview on Twin Support Vector Machines. Artificial Intelligence Review, 2012 (DOI:10.1007/s10462-012-9336-0).
- [18] Junzhao Yu, Shifei Ding, Fengxiang Jin, Huajuan Huang, Youzhen Han. Twin Support Vector Machines Based on Rough Sets. International Journal of Digital Content Technology and its Applications, 2012, 6(20):493-500.
- [19] Chen Jing, Ji Guangrong. Weighted least squares twin support vector machines for pattern classification. 2010 The 2nd International Conference on Computer and Automation Engineering, Singapore: [s.n.], 2010, 2:242-246.
- [20] Qi Zhiqian, Tian Yingjie, Shi Yong. Robust twin support vector machine for pattern classification. Pattern Recognition. 2013, 46(1):305-316.
- [21] Xinsheng Zhang, Xinbo Gao, Ying Wang. MCs Detection with Combined Image Features and Twin Support Vector Machines. Journal of Computers, 2009, 4(3), 215-221.
- [22] Huajuan Huang, Shifei Ding, Fengxiang Jin, Junzhao Yu, Youzhen Han. A Novel Granular Support Vector Machine Based on Mixed Kernel Function. International Journal of Digital Content Technology and its Applications, 2012, 6(20):484-492.
- [23] Wang Guosheng. Properties and Construction Methods of Kernel in Support Vector Machine. Computer Science, 2006, 33(6): 172-174.



computing, feature learning et al.

Fulin Wu is currently a graduate student now studying in School of Computer Science and Technology, China University of Mining and Technology, and his supervisor is Prof. Shifei Ding. He received his B.Sc. degree in computer science from China University of Mining and Technology in 2012. His research interests include cloud



Shifei Ding is a professor and Ph.D. supervisor at China University of Mining and Technology. He received his B.Sc., M.Sc. degree in mathematics and computer science in 1987, 1998 respectively from Qufu Normal University, received Ph.D. degree in computer science from Shandong University of Science and Technology in 2004. His research interests include

intelligent information processing, pattern recognition, machine learning, data mining, and granular computing et al. He has published 3 books and more than 150 research papers in journals and international conferences.