Invasive Weed Optimization Algorithm for Optimizating the Parameters of Mixed Kernel Twin Support Vector Machines

Huajuan Huang
School of Computer Science and Technology, China University of Mining and Technology, Xuzhou, China, 221116
Email: hhj-025@163.com

Shifei Ding
School of Computer Science and Technology, China University of Mining and Technology, Xuzhou, China, 221116
Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, 100190
Email: dingsf@cumt.edu.cn

Hong Zhu, Xinzheng Xu
School of Computer Science and Technology, China University of Mining and Technology, Xuzhou, China, 221116
Email: zhuhongwin@126.com; xuxinzh@163.com

***Abstract*— How to select the suitable parameters and kernel model is a very important problem for Twin Support Vector Machines (TSVMs). In order to solve this problem, one solving algorithm called Invasive Weed Optimization Algorithm for Optimizating the Parameters of Mixed Kernel Twin Support Vector Machines (IWO-MKTSVMs) is proposed in this paper. Firstly, introducing the mixed kernel, the twin support vector machines based on mixed kernel is constructed. This strategy is a good way to solve the kernel model selection. In order to solve the parameters selection problem which contain TSVMs parameters and mixed kernel model parameters, Invasive Weed Optimization Algorithm (IWO) is introduced. IWO is an optimization algorithm who has strong robustness and good global searching ability. Finally, compared with the classical TSVMs, the experimental results show that IWO-MKTSVMs have higher classification accuracy.**

***Index Terms*— Mixed kernel, Invasive weed optimization algorithm, Parameter optimization, Twin support vector machines**

I. INTRODUCTION

Support Vector Machines (SVM) is known as a new generation learning system based on statistical learning theory [1-4]. Because of its profound mathematical theory, SVM has played excellent performance on many real-world predictive data mining applications such as text categorization, medical and biological information analysis [5-7],etc.

One of the main challenges for the traditional SVM is the high computational complexity. The training cost of $O(n^3)$, where n is the total size of the training data, is too expensive. In order to improve the computational speed, Jayadeva et al. [8] proposed a new machine learning method called Twin Support Vector Machines (TSVMs) for the binary classification in the spirit of proximal SVM [9-10] in 2007. TSVMs would generate two non-parallel planes, such that each plane is closer to one of the two classes and is as far as possible from the other. In TSVMs, a pair of smaller sized quadratic programming (QP) problems is solved, whereas SVM solves a single QP problem. Furthermore, in SVM, the QP problem has all data points in the constraints, but in TSVMs they are distributed in the sense that patterns of class -1 give the constraints of the QP used to determine the hyperplane for class 1, and vice-versa. This strategy of solving two smaller sized QP problems, rather than one larger QP problem, makes the computational speed of TSVMs approximately 4 times faster than the traditional SVM. Because of its excellent performance, TSVMs has been applied to many areas such as speaker recognition [11], medical detection [12-14], etc. At present, many improved TSVMs algorithms have been proposed. For example, in 2010, M.Arun Kumar et al. [15] brought the prior knowledge into TSVMs and least square TSVMs and then got two improved algorithms. Experimental results showed the proposed algorithms were effective. In 2011, Qiaolin Yu et al.[16] adding the regularization method into the TSVMs model, proposed the TSVMs model based on regularization method. This method ensured that the proposed model was the strongly convex programming problem. In 2012, Yitian Xu et al.[17] proposed a twin multi-class classification support vector machine. Experimental results demonstrated the proposed algorithm was stable and effective.

As a new machine learning methods, there are still many places needing to be perfect for TSVMs. Specially,

Manuscript received October 21, 2012; revised December 1, 2012; accepted January 15, 2013
Corresponding author: Shifei Ding

doi:10.4304/jcp.8.8.2077-2084

the learning performance and generalization ability of TSVMs is very dependent on its parameters and kernel model selection. If the choice is reasonable, it will be very difficult to approach superiorly. However, the current research on this aspect is very little. At present, the kernel model selection adopts the random or experimental method. These methods are blindness and time consuming. For the parameters selection, the grid search method is commonly used. However, the search time of this method is too long, especially in dealing with the large dataset. In order to solve this problem, one solving algorithm called Invasive Weed Optimization Algorithm for Optimizating the Parameters of Mixed Kernel Twin Support Vector Machines (IWO-MKTSVMs) is proposed in this paper. Firstly, in view of the blindness of the kernel model selection for TSVMs, one kernel function with good generalization ability and the other kernel function with good learning ability is combined, formed a mixed kernel model with the more excellent performance. Secondly, because of the limitation of the traditional selection method for TSVMs, we use Invasive Weed Optimization (IWO) algorithm which has fast global searching ability to select the TSVMs parameters and the mixed kernel parameters, so that we would obtain the optimal parameters combination. Finally, the experimental results show the effectiveness and stability of the proposed method.

The paper is organized as follows: In section 2, we briefly introduce the basic theory of TSVMs and the analysis of its parameters. In section 3, IWO-MKTSVMs algorithm is detailed introduced and analyzed. Computational comparisons on UCI datasets are done in section 4 and section 5 gives concluding remarks.

II. Twin Support Vector Machines and Its Parameters

A. Twin Support Vector Machines

Consider a binary classification problem of classifying m_1 data points belonging to class +1 and m_2 data points belonging to class -1. Then let matrix A in $R^{m_1 \times n}$ represent the data points of class +1 while matrix B in $R^{m_2 \times n}$ represent the data points of class -1. Two nonparallel hyper-planes of the linear TSVMs can be expressed as follows.

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

The target of TSVMs is to generate the above two nonparallel hyper-planes in the n-dimensional real space R^n, such that each plane is closer to one of the two classes and is as far as possible from the other. A new sample point is assigned to class +1 or -1 depending upon its proximity to the two nonparallel hyper-planes. The linear classifiers are obtained by solving the following optimization problems.

$$\begin{aligned} &\min_{w^{(1)},b^{(1)},\xi^{(2)}} \frac{1}{2}\left\|Aw^{(1)}+e_1b^{(1)}\right\|^2 + c_1e_2^T\xi^{(2)} \\ &s.t. \quad -(Bw^{(1)}+e_2b^{(1)}) \ge e_2 - \xi^{(2)}, \\ &\qquad \xi^{(2)} \ge 0. \end{aligned} \tag{2}$$

$$\begin{aligned} &\min_{w^{(2)},b^{(2)},\xi^{(1)}} \frac{1}{2}\left\|Bw^{(2)}+e_2b^{(2)}\right\|^2 + c_2e_1^T\xi^{(1)} \\ &s.t. \quad (Aw^{(2)}+e_1b^{(2)}) \ge e_1 - \xi^{(1)}, \\ &\qquad \xi^{(1)} \ge 0. \end{aligned} \tag{3}$$

where c_1 and c_2 are penalty parameters, $\xi^{(1)}$ and $\xi^{(2)}$ are slack vectors, e_1 and e_2 are the vectors of ones of appropriate dimensions. $x_j^{(i)}$ represents the j th sample of the i th class. Introducing the Lagrange variables α and β, the dual problems of (2) and (3) can be expressed as follows:

$$\begin{aligned} &\max_{\alpha} \ e_2^T\alpha - \frac{1}{2}\alpha^T G(H^TH)^{-1}G^T\alpha \\ &s.t. \quad 0 \le \alpha \le c_1e_2 \end{aligned} \tag{4}$$

$$\begin{aligned} &\max_{\beta} \ e_1^T\beta - \frac{1}{2}\beta^T H(G^TG)^{-1}H^T\beta \\ &s.t. \quad 0 \le \beta \le c_2e_1, \end{aligned} \tag{5}$$

where, $H = [A \quad e_1]$, $G = [B \quad e_2]$. Defining $u_i = [(w^{(i)})^T \quad b^{(i)}]^T$, $i = 1,2$ the solution becomes:

$$u_1 = -(H^TH)^{-1}G^T\alpha\,, \quad u_2 = (G^TG)^{-1}H^T\beta \tag{6}$$

To judge a new sample belonging to which class, we should find this sample is closer to which class. We can calculate the distance of a sample from a class by (7).

$$f(x) = \arg\min_i (d_i(x)), \tag{7}$$

where, $$d_i(x) = \frac{\left|x^Tw^{(i)} + b^{(i)}\right|}{\left\|w^{(i)}\right\|_2}, \quad i = 1,2\,. \tag{8}$$

For the nonlinear case, the two nonparallel hyperplanes of TSVMs based on kernel can be expressed as follows:

$$K(x^T, C^T)w^{(1)} + b^{(1)} = 0,\ K(x^T, C^T)w^{(2)} + b^{(2)} = 0$$

where, $C = [A^T, B^T]^T$. So the optimization problem of nonlinear TSVMs can be expressed as follows.

$$\begin{aligned} &\min_{w^{(1)},b^{(1)},\xi^{(2)}} \frac{1}{2}\left\|K(A,C^T)w^{(1)}+e_1b^{(1)}\right\|^2 + c_1e_2^T\xi^{(2)} \\ &s.t. \quad -(K(B,C^T)w^{(1)}+e_2b^{(1)}) \ge e_2 - \xi^{(2)}, \\ &\qquad \xi^{(2)} \ge 0. \end{aligned} \tag{9}$$

$$\begin{aligned} &\min_{w^{(1)},b^{(1)},\xi^{(2)}} \frac{1}{2}\left\|K(B,C^T)w^{(2)}+e_2b^{(1)}\right\|^2 + c_2e_1^T\xi^{(1)} \\ &s.t. \quad (K(A,C^T)w^{(2)}+e_1b^{(2)}) \ge e_1 - \xi^{(1)}, \\ &\qquad \xi^{(1)} \ge 0. \end{aligned} \tag{10}$$

According to the lagrange theorem, the dual problems of (9) and (10) can be expressed by (11) and (12).

$$\begin{aligned} &\max_{\alpha} \ e^T\alpha - \frac{1}{2}\alpha^T R(S^TS)^{-1}R^T\alpha \\ &s.t. \quad 0 \le \alpha \le c_1e_2 \end{aligned} \tag{11}$$

$$\max_{\alpha} \ e^T\beta - \frac{1}{2}\beta^T S(R^TR)^{-1}S^T\beta$$

$$s.t. \quad 0 \le \beta \le c_2 e_1 \tag{12}$$

where, $S = [K(A,C) \quad e_1]$, $R = [K(B,C) \quad e_2]$. Defining $v_i = [(w^{(i)})^T \quad b^{(i)}]^T$, $i = 1,2$ the solution becomes:

$$v_1 = -(S^T S)^{-1} R^T \alpha$$

$$v_2 = (R^T R)^{-1} S^T \beta \tag{13}$$

B. Analysis the Penalty Parameters of TSVMs

The role of penalty parameters c_1 and c_2 is to adjust the ratio between the confidence range with the experience risk in the defining feature, so that the generalization ability of TSVMs can achieve the best state. The values of c_1 and c_2 smaller expresses the punishment on empirical error smaller. Do it this way, the complexity of TSVMs is smaller, but its fault tolerant ability is worse. The values of c_1 and c_2 are greater, the data fitting degree is higher, but its generalization capacity will be reduced. From the above analysis, we can know that the parameters selection is very important for TSVMs.

III. Invasive Weed Optimization Algorithm for Optimizating The Parameters of Mixed Kernel Twin Support Vector Machines

A. Kernel Function

Similar to SVM, by introducing the kernel function, TSVMs can achieve the linear classification in the dimensional feature for the nonlinear problems. Therefore, kernel function takes an important role in the nonlinear TSVMs. At present, the most commonly used kernel functions in TSVMs are as follows:

1) The linear kernel function:

$$K(x, x_i) = x, x_i \tag{14}$$

2) The polynomial kernel function:

$$K(x, x_i) = (\gamma(x, x_i) + 1)^d, \gamma > 0 \tag{15}$$

3) The gauss kernel function:

$$K(x, x_i) = \exp(-\frac{\|x - x_i\|^2}{2\sigma^2}) \tag{16}$$

4) The sigmoid kernel function:

$$K(x, x_i) = \tanh(p_1(x, x_i) + p_2) \tag{17}$$

The choice of kernel function is a critical problem in the practical application. This is because that the learning ability of kernel function will directly affect the quality of kernel model performance realization. Kernel functions have many characteristics. Summed them up, kernel functions can be divided into two types, i.e., global kernel functions and local kernel function. For the global kernel function, its generalization ability is strong when its learning ability is weak. On the contrary, for the local kernel function, it has strong learning ability but its generalization ability is weak. In view of the respective characteristic of the global and local kernel function, if the two type of kernel functions are mixed into a hybrid kernel function, which will be able to achieve the good classification performance. Based on the above ideas, we will construct a mixed kernel function as follows.

As we know, the most used kernel function in TSVMs is the Gauss kernel which is a typical local kernel function. For the Gauss kernel function, the sketch map of the testing point 0.1 is shown as figure 1. When the values of σ^2 are 0.1, 0.2, 0.3, 0.4, 0.5 respectively.

From figure 1 we can see that the Gauss function has good learning ability because of only having a role for the near test point, but its generalization ability is weak.

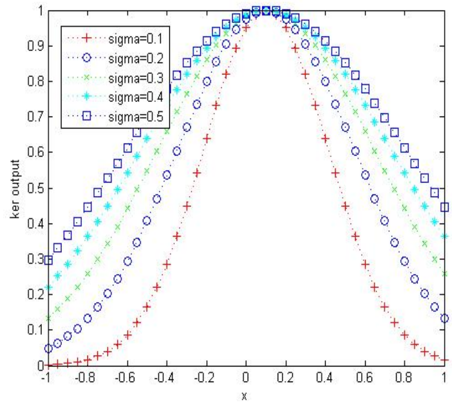

Figure 1. The curve of Gaussian kernel in test point 0.1

The polynomial kernel is a typical global kernel function. Compared with the local kernel function, the learning ability of global kernel function is weak, but it has good generalization ability. For polynomial kernel function, the sketch map of the testing point 0.1 is shown as figure 2. when the values of q are 1, 2, 3, 4, 5 respectively.

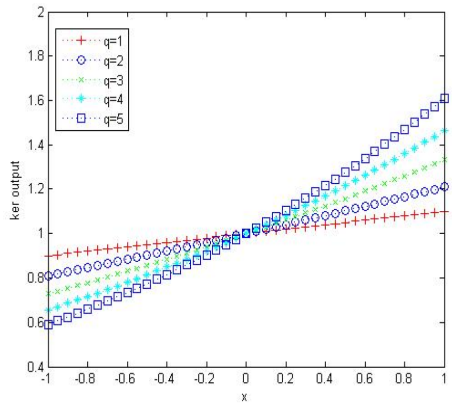

Figure 2. The curve of polynomial kernel in test point 0.1

But the learning ability in the test point is not obvious, which means its learning ability is not only strong.

B. Construction Mixed Kernel Function

Based on the above analysis, if the Gauss kernel function and the polynomial kernel function is mixed to generate a new mixed kernel function, which can have better learning ability and better generalization ability.

Theorem 1 The training samples are linear separability in the kernel space when $rank(K) = n$, where $K = (k(x_i, x_j))_{l \times l}$ is Gram matrix, $k(x_i, x_j)$ is the kernel function.

For the Gauss kernel function, Gram matrix K is a strong diagonal matrix when $\sigma \to 0$. Therefore, K is full rank. From theorem 1, we can see that the training samples are linear separability when the kernel function is the Gauss kernel function.

Theorem 2 The linear mixed function

$$\tilde{k}(x_i, x_j) = a \cdot \exp(-\frac{\|x_i - x_j\|^2}{2\sigma^2}) + (1-a)\cdot k(x_i, x_j),$$

($0 < a < 1$) which contains Gauss kernel function $K(x_i, x_j) = \exp(-\frac{\|x_i - x_j\|^2}{2\sigma^2})$ and any kernel function $k(x_i, x_j)$ is a new kernel function.

Proof According to Mecer theorem, the Gram matrix K of any kernel function $k(x_i, x_j)$ is symmetric and positive semi-definite. So there is orthogonal matrix U,

let $K = U\begin{bmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \lambda_k & \\ & & & 0 \end{bmatrix}U^{-1}$, where $\lambda_i > 0$. Therefore,

when $\sigma \to 0$, there is

$$\tilde{k} \approx a \cdot I + (1-a)\cdot K = U\begin{bmatrix} a & & & \\ & a & & \\ & & \ddots & \\ & & & a \end{bmatrix}U^{-1}$$

$$+U\begin{bmatrix} (1-a)\lambda_1 & & & \\ & \ddots & & \\ & & (1-a)\lambda_k & \\ & & & 0 \end{bmatrix}U^{-1}$$

$$=U\begin{bmatrix} a+(1-a)\cdot\lambda_1 & & & & \\ & \ddots & & & \\ & & a+(1-a)\cdot\lambda_k & & \\ & & & \ddots & \\ & & & & a \end{bmatrix}U^{-1}$$

So we can know that $\tilde{k}$ is full rank. According to theorem 1, $\tilde{k}$ is a kind of kernel function can linearly separate any training samples.

Based on the above analysis, we can obtain the mixed function as follows:

$$\tilde{k}(x_i, x_j) = a \cdot \exp(-\frac{\|x_i - x_j\|^2}{2\sigma^2}) + (1-a)[(x_i \cdot x_j) + 1]^q$$

According to theorem 2, the above formula is a kernel function. For the mixed kernel function, the sketch map of the testing point 0.1 when σ^2=0.1, $q = 3$ is shown as figure 3.

From the Figure 3, we can see that the mixed kernel function makes good use of the properties of global and local kernels. It not only has the strong learning capability but also has strong generalization ability.

C. Analysis the Parameters of Mixed Kernel Function

After introducing the mixed kernel, TSVMs have added three adjustable parameters which contain the weight of mixed kernel function a, the Gauss kernel function parameter σ and the polynomial kernel parameter q.

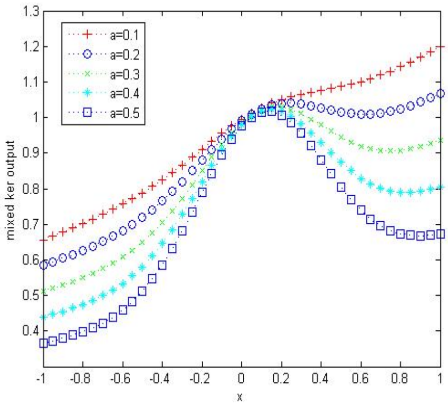

Figure 3. The curve of mixed kernel in test point 0.1

According to the properties of kernel function, the value of a is between 0 : 1. The mixed kernel function is closer to the polynomial kernel function when $a \to 0$. On the contrary, the mixed kernel function is closer to the Gauss kernel function when $a \to 1$. Therefore, it is very important to select a. If the choice is not appropriate, it may make the performance of mixed kernel function below the single one, thus losing the advantage of mixed kernel function. σ and q are the kernel parameters, which also take important role in the performance of mixed kernel function. At present, there are two selection methods of kernel parameters. One is the random method and the other is cross validation method. The random method is that the kernel parameters are randomly given and then the value of kernel parameters is constantly adjusted until getting a satisfactory precision. In view of lack of adequate theoretical basis, the random method has certain blindness. The cross validation method tests a range of kernel parameters individually to find the optimal value using traversal approach. Generally, this method can find the best values, but its time complexity is relatively high.

After the above analysis, in this paper, Invasive Weed Optimization (IWO) algorithm which has fast global searching ability is used to select the TSVMs parameters and the mixed kernel parameters.

D. Invasive Weed Optimization

In 2006, a novel stochastic optimization model, invasive weed optimization (IWO) algorithm [18], was proposed by Mehrabian and Lucas, which is inspired from a common phenomenon in agriculture: colonization of invasive weeds. Not only it has the robustness, but also it is easy to understand and program. So far, it has been applied in many engineering fields [19-20].

In the classical IWO, weeds represent the feasible solutions of problems and population is the set of all weeds. A finite number of weeds are being dispread over the search area. Every weed produces new weeds depending on its fitness. The generated weeds are randomly distributed over the search space by normally distributed random numbers with a mean equal to zero. This process continues until maximum number of weeds is reached. Only the weeds with better fitness can survive and produce seed, others are being eliminated. The process continues until maximum iterations are reached or hopefully the weed with best fitness is closest to optimal solution.

The process is addressed in details as follows:

Step 1: *Initialize a population*

A population of initial solutions is being dispread over the D dimensional search space with random positions.

Step 2: *Reproduction*

The higher the weed's fitness is, the more seeds it produces. The formula of weeds producing seeds is

$$weed_n = \frac{f - f_{\min}}{f_{\max} - f_{\min}}(s_{\max} - s_{\min}) + s_{\min} \tag{18}$$

where, f is the current weed's fitness. f_{max} and f_{min} respectively represent the maximum and the least fitness of the current population. s_{max} and s_{min} respectively represent the maximum and the least value of a weed.

Step 3: *Spatial dispersal*

The generated seeds are randomly distributed over the D dimensional search space by normally distributed random numbers with a mean equal to zero, but with a varying variance. This ensures that seeds will be randomly distributed so that they abide near to the parent plant. However, standard deviation (σ) of the random function will be reduced from a previously defined initial value (σ_{init}) to a final value (σ_{final}) in every generation. In simulations, a nonlinear alteration has shown satisfactory performance, given as follows

$$\sigma_{cur} = \frac{(iter_{\max} - iter)^n}{(iter_{\max})^n}(\sigma_{init} - \sigma_{final}) + \sigma_{final} \tag{19}$$

Where, $iter_{max}$ is the maximum number of iterations, σ_{cur} is the standard deviation at the present time step and n is the nonlinear modulation index. Generally, n is set to 3.

Step 4: *Competitive exclusion*

After passing some iteration, the number of weeds in a colony will reach its maximum (P_MAX) by fast reproduction. At this time, each weed is allowed to produce seeds. The produced seeds are then allowed to spread over the search area. When all seeds have found their position in the search area, they are ranked together with their parents (as a colony of weeds). Next, weeds with lower fitness are eliminated to reach the maximum allowable population in a colony. In this way, weeds and seeds are ranked together and the ones with better fitness survive and are allowed to replicate. The population control mechanism also is applied to their offspring to the end of a given run, realizing competitive exclusion.

E. The Algorithm Steps of IWO-MKTSVMs

The accuracy in the sense of CV is used for the fitness of IWO. So the algorithm steps of IWO-MKTSVMs are as follows:

Step1: Select the training dataset and the testing dataset.

Step2: Preprocessing the dataset.

Step3: Constructe the mixed kernel function.

Step4: Select the optimal parameters using IWO algorithm.

Step5: Train the mixed kernel TSVMs using the optimal parameters.

Step6: Predict the testing dataset.

Step7: Output the classification accuracy.

IV. The Experiment Results and Analysis

In order to verify the efficiency of IWO-MKTSVMs, meanwhile, in order to compare the performance of three algorithms, that is, SVM, TSVMs and IWO-MKTSVMs, we conduct experiments on seven benchmark datasets from the UCI machine learning repository. The environments of all experiments are in Intel (R) Core (TM) 2Duo CUP E4500, 2G memory and MATLAB 7.11.0. The parameter values of IWO are as follows: $D=5$, $P\_MAX=30$, $s_{\max}=5$, $s_{\min}=1$, $n=3$, $\sigma_{init}=[1,0.1,1,1,1]$, $\sigma_{final}=[0.1,0.1,0.1,0.1,0.1]$. In IWO algorithm, the accuracy in the sense of CV is used for the fitness of IWO. Therefore, the fitness value is closer to 100, the obtained parameters is closer to the optimal value. The experiment results of IWO-MKTSVMs are shown as table 1. Furthermore, the comparisons of IWO-MKTSVMs and other algorithms are shown as table 2. In order to more objectively test the performance of each algorithm, we test each dataset 20 times independently. And the values of table 1 and table 2 are the average values. Figure 4 and figure 5 are the fitness curves of IWO searching the optimal parameters for dealing with the Australian dataset and Breast-cancer dataset respectively. Figure 6 represents the classification results on seven UCI dataset by three algorithms.

From table 1, we can see that the training accuracy and testing accuracy of IWO-MKTSVMs is relatively high. Meanwhile, table 1 lists the optimal parameters using IWO for searching. Table 2 is the testing accuracy comparisons of IWO-MKTSVMs , TSVMs and SVM. From table 2, we know that the classification results of IWO-MKTSVMs are better than the other algorithms. Figure 4 and figure 5 shows that the optimization ability of IWO is very strong.

V. Conclusion And Furture Work

In order to solve the problems of selecting the parameters and kernel model for TSVMs, one solving algorithm called Invasive Weed Optimization Algorithm for Optimizating the Parameters of Mixed Kernel Twin Support Vector Machines (IWO-MKTSVMs) is proposed in this paper.

Firstly, by introducing the mixed kernel function, we obtain a kind of kernel function with good performance, which can solve the problem of selecting kernel function in TSVMs. Secondly, in view of the good optimization ability of Invasive Weed Optimization (IWO) algorithm, it is used to optimize the parameters containing the TSVMs parameters and the mixed kernel parameters. Finally, the experimental results show the effectiveness and stability of the proposed method. How to further improve the performance of IWO is the next work.

TABLE1.

THE CLASSIFICATION RESULTS OF IWO-MKTSVMs

Dataset	The optimal parameter values					Training Accuracy(%)	Testing Accuracy(%)
	c_1	c_2	a	σ	q		
Australian	1.414	5.241	0.752	12.042	36.457	98.63 ± 1.21	87.65 ± 4.12
Breast-cancer	2.922	15.698	0.524	85.770	12.044	83.16 ± 5.28	69.12 ± 4.23
Heart	45.88	54.362	0.635	92.617	52.017	91.25 ± 6.35	84.22 ± 7.21
Pima	54.68	1.487	0.821	4.227	12.781	94.25 ± 4.22	82.09 ± 2.02
Votes	83.08	2.514	0.424	17.052	3.654	99.21 ± 0.21	96.23 ± 2.30
Sonar	6.241	8.695	0.832	2.044	60.140	94.28 ± 1.25	90.45 ± 4.25
CMC	62.47	69.214	0.781	36.451	96.012	87.854 ± 4.21	75.54 ± 7.25

TABLE 2.

TESTING ACCURACY COMPARISONS OF IWO-MKTSVMs AND OTHER ALGORITHMS

Dataset	IWO-MKTSVMs	TSVMs	SVM
Australian	87.65 ± 4.12	84.81 ± 2.15	85.51 ± 2.16
Breast-cancer	69.12 ± 4.23	64.42 ± 3.87	65.42 ± 4.53
Heart	84.22 ± 7.21	81.89 ± 4.31	82.22 ± 6.67
Pima	82.09 ± 2.02	73.70 ± 6.05	76.55 ± 2.40
Votes	96.23 ± 2.30	94.96 ± 4.24	95.85 ± 2.24
Sonar	90.45 ± 4.25	89.52 ± 3.37	88.91 ± 9.68
CMC	75.54 ± 7.25	73.50 ± 9.85	68.98 ± 2.17

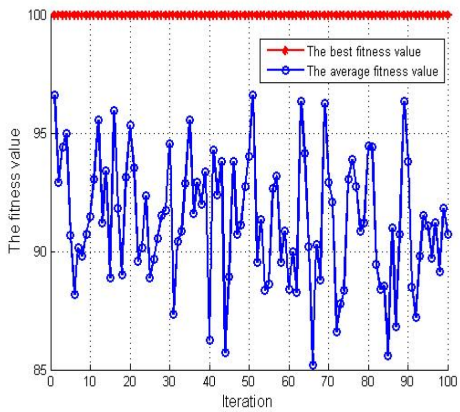

Figure 4. The fitness curves of IWO searching the optimal parameters for dealing with the Australian dataset

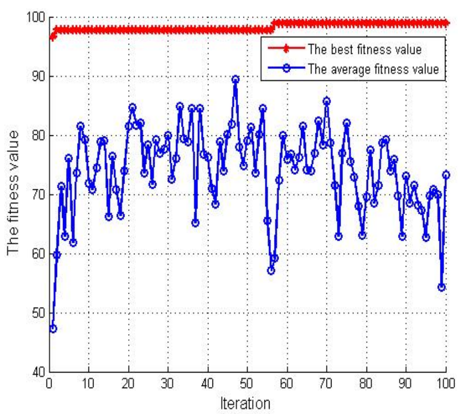

Figure 5. The fitness curves of IWO searching the optimal parameters for dealing with the Breast-cancer dataset

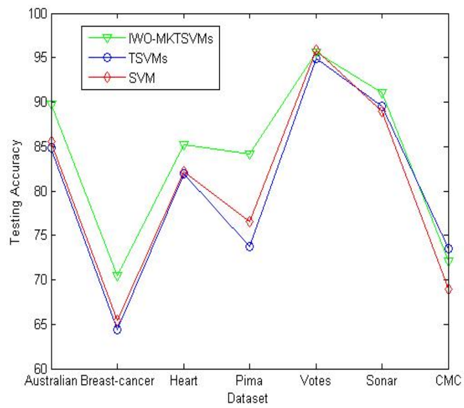

Figure 6. The classification results on seven UCI dataset by three algorithms

ACKNOWLEDGMENT

This work is supported by the National Key Basic Research Program of China (No. 2013CB329502), the National Natural Science Foundation of China (No.41074003), the Opening Foundation of the Key Laboratory of Intelligent Information Processing of Chinese Academy of Sciences (IIP2010-1).

REFERENCES

[1] S.F. Ding and B.J. Qi, "Research of granular support vector machine", *Artificial Intelligence Review*, 2012, vol. 38, no.1,

pp.1-7.
[2] V.N.Vapnik. The Nature of Statistical Learning *Theory. New York: Springer-Verlag*,1995.
[3] Rodrigo Moraes, Joao Francisco Valiati, Neto Gaviao, P Wilson. Document-level sentiment classification: An empirical comparison between SVM and ANN. *Expert Systems with Application*. 2013, vol. 40, no.2, pp.621-633.
[4] Jiaxin Wu. Efficient HIK SVM learning for Image Classification. *IEEE Transactions on Image Processing*. 2012, vol.21, no.10, pp.4442-4453.
[5] L. Hao, X.M. Tao, P.J. Xu and G..X. Qiu, "Classification of Bio-potential Surface Electrode based on FKCM and SVM", *Journal of Software*, 2011, vol. 6, no.5, pp.880-886.
[6] X.M. Wang, Y.Y. Liang, X.J. Zhao and Z.L. Wang, "Lip Aus Detection by Boost-SVM and Gabor", *Journal of Software*, 2012, vol. 7, no.9, pp.1968-1974.
[7] X.J. Li, K. Wang and L.L. Jiang, "The Application of AE Signal in Early Cracked Rotor Fault Diagnosis with PWND and SVM", *Journal of Software*, 2011, vol. 6, no.10, pp.1969-1976.
[8] Jayadeva, K. Reshma and S. Chandra, "Twin support vector machines for pattern classification", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, vol. 29, no.5, pp. 905-910.
[9] G. Fung and O. L. Mangasarian, "Proximal Support Vector Machine Classifiers", *In: Proc 7th ACM SIFKDD Intl Conf on Knowledge Discovery and Data Mining*, 2001, pp.77-86.
[10] O.L. Mangasarian and E.W. Wild, "Multisurface proximal support vector machine classification via generalized eigenvalues", *IEEE Transactions on Pattern Aalysis and Machine Intelligence*, 2006, vol. 28, no.1, pp.69-74.
[11] H.H. Cong, C.F. Yang and X.R. Pu, "Efficient Speaker Recognition based on Multi-class Twin Support Vector Machines and GMMs", *2008 IEEE Conference on Robotics, Automation and Mechatronics*, 2008, pp.348-352.
[12] X.S. Zhang, X.B. Gao and Y. Wang, "Twin Support Vector Machine For MCs Detection", *Journal of Electronics (China)*, 2009, vol. 26, no.3, pp.318-325.
[13] X.S. Zhang, "Boosting twin support vector machine approach for MCs detection", *Asia-Pacific Conference on Information Processing*, 2009, vol. 46, pp.149-152.
[14] X.S. Zhang and X.B. Gao, "MCs Detection Approach Using Bagging and Boosting Based Twin Support Vector Machine", *2009 IEEE International Conference on Systems, Man, and Cybernetics San Antonio, TX, USA*, 2009, pp.5000-5005.
[15] M. A. Kumar, R. Khemchandani, M. Gopal and S. Chandra, "Knowledge based Least Squares Twin support vector machines", *Information Sciences*, 2010, vol. 180, no.23, pp: 4606–4618.
[16] Q.L. Ye, C.X. Zhao and X.B. Chen, "A Feature Selection Method for TWSVM via a Regularization Technique", *Journal of Computer Research and Development*, 2011, vol. 48, no.6, pp:1029-1037.
[17] Y.T. Xu, R. Guo and L.S. Wang, "A Twin Multi-Class Classification Support Vetor Machine", *Cogn Comput*, 2012.(DOI 10.1007/s12559-012-9179-7).
[18] A.R. Mehrabian, C. Lucas, "A novel numerical optimization algorithm inspired from weed colonization", *Ecological Informatics*, 2006, vol. 1, no.4, pp:355-366.
[19] G.R. Gourab, D. Swagatam, C. Prithwish and N.S. Ponnuthurai, "Design of Non-Uniform Circular Antenna Arrays Using a Modified Invasive Weed Optimization Algorithm", *IEEE Transactions on antennas and propagation*, vol.59, no.1, pp.110-118.
[20] F.M. Monavar and N. Komjani, "Bandwidth enhancement of microstrip patch antenna using jerusalem cross-shaped frequency selective surfaces by invasive weed optimization approach", *Progress In Electromagnetics Research*, vol.121,pp.103-120.

Huajuan Huang, born in 1984. Received her B.Sc.degree and M.Sc. degree in applied computer Technology from Guangxi University for Nationalities, Guangxi, China, in 2006 and 2009 respectively. Since 2011, she has been a Ph.D. degree candidate in applied computer Technology from the China University of Mining and Technology, Xuzhou, China. Her current research interests include data mining, parrern recognition and computational intelligence.

Shifei Ding received his bachelor's degree and master's degree from Qufu Normal University in 1987 and 1998 respectively. He received his Ph.D degree from Shandong University of Science and Technology in 2004. He received postdoctoral degree from Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences in 2006. And now, he works in China University of Mining and Technology as a professor and Ph.D supervisor. 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 80 research papers in journals and international conferences. Prof. Ding is a senior member of China Computer Federation (CCF), and China Association for Artificial Intelligence (CAAI). He is a member of professional committee of distributed intelligence and knowledge engineering, CAAI, professional committee of machine learning, CAAI, and professional committee of rough set and soft computing, CAAI. He acts as an editor for Journal of Convergence Information Technology (JCIT), International Journal of Digital Content Technology and its Applications (JDCTA). Meanwhile, he is a reviewer for Journal of Information Science (JIS), Information Sciences (INS), Computational Statistics and Data Analysis (CSTA), IEEE Transactions on Fuzzy Systems(IEEE TFS), Applied Soft Computing (ASOC), Computational Statistics and Data Analysis (CSDA), International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI) et al.

Hong Zhu is an associate professor at Xuzhou Medical College. Since 2009, she has been a Ph.D. candidate in computer science and technology from the China University of Mining and Technology, and her supervisor is Prof. Shifei Ding. She received her M.Sc. degree in computer science and technology from China University of Mining and Technology in 2007.Her research interests includes granule computing, clustering, parallel computing et al. Email:zhuhongwin@126.com

Xinzheng Xu received his PH.D degree in Computer Application Technology from China University of Mining and Technology in 2012, and his MS degree in Computer Application Technology from Xiamen University in 2005. He is currently an associate professor at School of Computer Science and Technology, China University of Mining and Technology. He is a member of China Computer Federation, and China Association for Artificial Intelligence. His research interests include intelligent information processing, pattern recognition, machine learning, and granular computing et al.