Learning the Parameters for Least Squares Support Vector Machine

Shuxia Lu, Xiaoxue Fan, Lisha Hu
Key Lab. of Machine Learning and Computational Intelligence,
College of Mathematics and Computer Science, Hebei University,
Baoding, China

***Abstract*—the regularization parameter and kernel parameter play important roles in the performance of the least squares support vector machine (LS-SVM). Aimed at optimizing the LS-SVM's parameters, a fast method based on distance is presented. The method is by way of calculating the various types of distances in the feature space to determine the optimal kernel parameter. Since the method only needs to calculate some simple mathematical formulas, and avoids training the corresponding LS-SVM classifiers, the method can greatly reduce the training time. Experiment results show that the proposed method can improve the training speed.**

Keywords-LS-SVM; kernel parameter; distance

I. INTRODUCTION

Support Vector Machine (SVM) was developed by Vapnik et al. in the last decade of the 20th century [1]. SVM is based on statistical learning theory with excellent learning performance, so it becomes a research focus on machine learning session and has been successfully applied in many areas. LS-SVM was introduced by Suykens and Vandewalle [2]. LS-SVM is a reformulation of the standard SVM [3]. LS-SVM simplifies the solution process of standard SVM in a great extent by substituting the inequality constraints by equality counterparts. Consequently, the decision function can be obtained by solving a group of linear equalities rather than quadratic program. LS-SVM has been successfully used in the fields of pattern recognition and function approximation problems and achieves good results [4]. Its performance depends on the selection of parameters values like other learning algorithms. How to effectively and quickly choose the optimal parameters has become one of the important questions of improving the ability of LS-SVM learning and generalization.

The existing classical used parameters selection methods for LS-SVM are grid search method [5], generalization error estimation method [6], inter-cluster distance method [7] and some methods based on heuristic searches including genetic algorithms [8], ant colony algorithm [9], and particle swarm optimization algorithms [10]. Grid search is a conventional method to deal with discontinuous problems. However, when there are more than two parameters, it may become time consuming and intractable. So Grid search does not perform well in practice. These heuristic algorithms usually fall into the population prematurely when solving complex optimization problems and thus obtain results with low precision. In this paper, a novel algorithm of parameter selection is proposed based on the distance. Not only the distance between vectors in feature space but its own center is considered in the proposed method, we account the inter-cluster distance. The experiment results show that the proposed method costs much less computation time than using a grid search method to train an LS-SVM.

In the following, an introduction of LS-SVM algorithm is presented in Section Ⅱ. The Novel LS-SVM Parameter Selection Algorithm is explained in SectionⅢ. In Section Ⅳ the experimental results are discussed. In Section Ⅴ we draw a general conclusion.

II. LEAST SQUARES SUPPORT VECTOR MACHINE

Like SVM, LS-SVM algorithm solves a convex problem. It has been shown that the generalization performance of LS-SVM can be comparable to that of SVM [11]. In addition, a linear problem is resolved instead of a quadratic programming (QP) problem. So the training algorithm of LS-SVM is more simply than SVM.

A. LS-SVM classification algorithm is described below

Let us consider a given training data set $T=\{(x_1,y_1),\cdots,(x_n,y_n)\}$, $x_i\in R^d$, $y_i\in\{1,-1\}$, $i=1,\cdots,n$, where x_i is the input vector, y_i is the label of x_i. Nonlinear LS-SVM classifiers use the kernel trick to produce nonlinear boundaries. That is introducing $\phi(x):R^d\to H$, where $\phi(x)$ is a mapping function that maps the sample from input space to a higher dimension feature space. An optimal hyperplane is computed in a feature space to construct LS-SVM model.

In LS-SVM, the following optimization problem is formulated

$$\min_{w,b,\xi} J(w,\xi)=\frac{1}{2}\|w\|^2+\frac{1}{2}\gamma\sum_{i=1}^{n}\xi_i^2$$
$$subject\ \ to\ \ y_i\left[w^T\phi(x_i)+b\right]=1-\xi_i,\ i=1,\cdots,n \tag{1}$$

Where the first term corresponds to the LS-SVM generalization performance and the second term represents the accuracy of the model. $w\in H$ is weight vector and $\xi_i\in R$ is

the error committed when approximating the i^{th} sample. $b \in R$ is bias term and γ is a regularization factor. The Lagrangian of problem (1) can be defined as follows:

$$L(w,b,\xi,\alpha)=\frac{1}{2}\|w\|^2+\frac{1}{2}\gamma\sum_{i=1}^{n}\xi_i^2-\sum_{i=1}^{n}\alpha_i\left\{y_i\left[w^T\phi(x_i)+b\right]-1+\xi_i\right\} \quad (2)$$

Where $\alpha_i(i=1,2,\cdots,n)$ are the Lagrange multipliers. The KKT conditions can be expressed by

$$\frac{\partial L}{\partial w}=0\Rightarrow w=\sum_{i=1}^{n}\alpha_i y_i\phi(x_i), \quad (3)$$

$$\frac{\partial L}{\partial b}=0\Rightarrow \sum_{i=1}^{n}\alpha_i y_i=0, \quad (4)$$

$$\frac{\partial L}{\partial \xi_i}=0\Rightarrow \alpha_i=\gamma\xi_i, \quad (5)$$

$$\frac{\partial L}{\partial \alpha_i}=0\Rightarrow y_i\left[w^T\phi(x_i)+b\right]+\xi_i-1=0. \quad (6)$$

Equation (3) to (6) can be translated into the following matrix equation:

$$\begin{bmatrix} I & 0 & 0 & -Z^T \\ 0 & 0 & 0 & -Y^T \\ 0 & 0 & \gamma I & -I \\ Z & Y & I & 0 \end{bmatrix}\begin{bmatrix} w \\ b \\ \xi \\ \alpha \end{bmatrix}=\begin{bmatrix} 0 \\ 0 \\ 0 \\ L_n \end{bmatrix} \quad (7)$$

Where I denotes a $n\times n$ identity matrix, $Z=\left[\phi(x_1)^T y_1,\cdots,\phi(x_n)^T y_n\right]^T$, $Y=[y_1,\cdots,y_n]^T$, $\xi=[\xi_1,\cdots,\xi_n]^T$, $\alpha=[\alpha_1,\cdots,\alpha_n]^T$, $L_n=[1,\cdots,1]^T$.

By eliminating of variables w, ξ, then (7) would be expressed as the following linear equations:

$$\begin{bmatrix} 0 & Y^T \\ Y & \Omega+\frac{1}{\gamma}I \end{bmatrix}\begin{bmatrix} b \\ \alpha \end{bmatrix}=\begin{bmatrix} 0 \\ L_n \end{bmatrix} \quad (8)$$

Where Ω is a symmetric matrix $\Omega=ZZ^T=\left[\Omega_{ij}\right]_{n\times n}$. According to Mercer's condition, we have $\Omega_{ij}=y_i y_j\phi(x_i)^T\phi(x_j)=y_i y_j K(x_i,x_j)$, $i,j=1,2,\cdots n$. Where $K(x_i,x_j)$ is a kernel function and the Mercer's condition has been applied in [12].

Thus (8) can be translated into the following linear equations:

$$\begin{bmatrix} 0 & y_1 & \cdots & y_n \\ y_1 & y_1y_1K(x_1,x_1)+1/\gamma & \cdots & y_1y_nK(x_1,x_n) \\ \vdots & \vdots & \vdots & \vdots \\ y_n & y_ny_1K(x_n,x_1) & \cdots & y_ny_nK(x_n,x_n)+1/\gamma \end{bmatrix}\cdot\begin{bmatrix} b \\ \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}=\begin{bmatrix} 0 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \quad (9)$$

The final results of the LS-SVM model for classification can be obtained as

$$y(x)=sign\left[\sum_{i=1}^{n}\alpha_i y_i K(x,x_i)+b\right]. \quad (10)$$

Where α, b are the solutions of (9). Any kernel function which satisfies the Mercer's condition can be used in LS-SVM. The kernel function $K(x_i,x_j)$ can typically be linear, polynomial and RBF kernel functions.

B. Some Common Kernel Functions

- Linear kernel: $K(x_i,x_j)=(x_i\cdot x_j)$
- Polynomial kernel: $K(x_i,x_j)=((x_i\cdot x_j)+c)^d \quad (d=1,2,\cdots)$
- RBF kernel: $K(x_i,x_j)=\exp\left(-\|x_i-x_j\|^2/2\sigma^2\right)$

The linear kernel is a special case of RBF kernels [8]. When training data set is linearly inseparable, the training data can be mapped into a higher feature space used a non-linear kernel, so non-linear kernels are better than linear kernels in the process of date processing. The polynomial kernel requires more parameters to be chosen, thus in this paper, we choice the RBF kernel as the kernel function.

C. A novel method of choosing parameters for LS-SVM

Since RBF function is used as the kernel function in this paper. In the LS-SVM algorithm, the optimization of the regularization parameter γ and kernel width parameter σ is an important step in establishing an efficient and high-performance LS-SVM model. So these parameters need to be properly tuned to minimize the generalization error.

In this paper, we propose a novel LS-SVM parameters selection method based on distance. For given Parameters search interval, when training data set is nonlinear, training data would be mapped into a higher-dimensional feature space via a mapping function ϕ, and then the inter-cluster distance and the distance between the vector in feature space and its own center were calculated, the optimal parameter σ^* of the kernel function must minimize the distance that between the vector in feature space and its own center, while maximizing the mean distance between the vector in feature space and the other labeled-center. When the ratio of the two distances is the smallest, we determine that the parameter σ^* is the best. Then we train the corresponding LS-SVM classifier separately for each parameter combination $<\sigma^*,\gamma>$ with the optimal kernel parameter σ^*. The regularization parameter γ is one of the given values. We choose the parameter γ^* as the optimal regularization parameter when LS-SVM classifier has the highest validation accuracy. Experiment results show that the proposed select parameters method can effectively shorten the training time and improve the training speed.

For the kernel function $K(x_i,x_j)$ with parameter σ, let the corresponding mapping is $\phi(x)$, that is, the input vector x is mapped to $\phi(x)$ in higher dimension feature space. So we have $K(x_i,x_j)=\phi(x_i)\cdot\phi(x_j)$.

Let us consider two training data sets $\{x_1, x_2, \cdots x_{n_1}\}$ and $\{x'_1, x'_2, \cdots, x'_{n_2}\}$, which are mapped to $\{\phi(x_1), \phi(x_2), \cdots \phi(x_{n_1})\}$ and $\{\phi(x'_1), \phi(x'_2), \cdots \phi(x'_{n_2})\}$ corresponding respectively. Let the centers of two training sets in feature space be m_ϕ and m'_ϕ, where $m_\phi = \frac{1}{n_1}\sum_{i=1}^{n_1}\phi(x_i)$, $m'_\phi = \frac{1}{n_2}\sum_{j=1}^{n_2}\phi(x'_j)$.

Let the average distance between the mapping vector $\phi(x_i)$ and its own center m_ϕ be $\delta_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} d(\phi(x_i), m_\phi)$, we can also note the average distance between the mapping vector $\phi(x'_j)$ and its own center m'_ϕ be $\delta_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} d(\phi(x'_j), m'_\phi)$. δ_1 and δ_2 represent the extent of compact for various classes. Let the sum distance between the mapping vector $\phi(x_i)$ and the other labeled center m'_ϕ be $\mu_1 = \sum_{i=1}^{n_1} d(\phi(x_i), m'_\phi)$, the sum distance between the mapping vector $\phi(x'_j)$ and the other labeled center m_ϕ be $\mu_2 = \sum_{j=1}^{l_2} d(\phi(x'_j), m_\phi)$.

The following objective function is defined to measure the classification reliability with different kernel parameters.

$$g(\sigma) = \frac{\delta_1 + \delta_2}{\frac{1}{n_1+n_2}(\mu_1+\mu_2)} = \frac{\frac{1}{n_1}\sum_{i=1}^{n_1} d(\phi(x_i), m_\phi) + \frac{1}{n_2}\sum_{j=1}^{n_2} d(\phi(x'_j), m'_\phi)}{\frac{1}{n_1+n_2}\left(\sum_{i=1}^{n_1} d(\phi(x_i), m'_\phi) + \sum_{j=1}^{n_2} d(\phi(x'_j), m_\phi)\right)} \quad (11)$$

Where the molecule represents the distance between the mapping vector and its own center, the denominator represents the inter-cluster distance.

From the Geometric point of view, the larger the inter-cluster distance, while the smaller the distance between the mapping vector and its own center, then the higher the classification reliability. Moreover, it is obvious that g is equivalent to the famous Fisher rule. So the g can indicate the class separation robustly. For any $\sigma \geq 0$, the optimal kernel parameter σ^* should be satisfied with $g(\sigma) \geq g(\sigma^*)$.

According to $K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$, we can get that the distance between two mapping vectors in the feature space can be evaluated by applying the L2-norm on the mapped data as the following equation:

$$\begin{aligned} d(\phi(x_i), \phi(x_j)) &= \sqrt{\|\phi(x_i) - \phi(x_j)\|^2} \\ &= \sqrt{\phi(x_i)\cdot\phi(x_i) - 2\phi(x_i)\cdot\phi(x_j) + \phi(x_j)\cdot\phi(x_j)} \\ &= \sqrt{K(x_i, x_i) - 2K(x_i, x_j) + K(x_j \cdot x_j)} \end{aligned} \quad (12)$$

Also the distance between two different class centers can be computed as

$$\begin{aligned} f = d(m_\phi, m'_\phi) &= \sqrt{\left\|\frac{1}{n_1}\sum_{i=1}^{n_1}\phi(x_i) - \frac{1}{n_2}\sum_{j=1}^{n_2}\phi(x'_j)\right\|^2} \\ &= \sqrt{\frac{1}{n_1^2}\sum_{i=1}^{n_1}\sum_{k=1}^{n_1}\phi(x_i)\cdot\phi(x_k) - \frac{2}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\phi(x_i)\cdot\phi(x'_j) + \frac{1}{n_2^2}\sum_{j=1}^{n_2}\sum_{p=1}^{n_2}\phi(x'_j)\cdot\phi(x'_p)} \\ &= \sqrt{\frac{1}{n_1^2}\sum_{i=1}^{n_1}\sum_{k=1}^{n_1}K(x_i, x_k) - \frac{2}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}K(x_i, x'_j) + \frac{1}{n_2^2}\sum_{j=1}^{n_2}\sum_{p=1}^{n_2}K(x'_j, x'_p)} \end{aligned} \quad (13)$$

The δ_i and the μ_j are defined by a kernel $K(x_i, x_j) = \phi(x_i)\cdot\phi(x_j)$, for example :

$$\begin{aligned} \delta_1 &= \frac{1}{n_1}\sum_{i=1}^{n_1} d(\phi(x_i), m_\phi) \\ &= \frac{1}{n_1}\sum_{i=1}^{n_1} d\left(\phi(x_i), \frac{1}{n_1}\sum_{j=1}^{n_1}\phi(x_j)\right) \\ &= \frac{1}{n_1}\sum_{i=1}^{n_1}\sqrt{\left\|\phi(x_i) - \frac{1}{n_1}\sum_{j=1}^{n_1}\phi(x_j)\right\|^2} \\ &= \frac{1}{n_1}\sum_{i=1}^{n_1}\sqrt{\phi(x_i)\cdot\phi(x_i) - \frac{2}{n_1}\phi(x_i)\sum_{j=1}^{n_1}\phi(x_j) + \frac{1}{n_1^2}\sum_{k=1}^{n_1}\sum_{j=1}^{n_1}\phi(x_k)\cdot\phi(x_j)} \\ &= \frac{1}{n_1}\sum_{i=1}^{n_1}\sqrt{K(x_i, x_i) - \frac{2}{n_1}\sum_{j=1}^{n_1}K(x_i, x_j) + \frac{2}{n_1^2}\sum_{k=1}^{n_1}\sum_{j=1}^{n_1}K(x_k, x_j)} \end{aligned} \quad (14)$$

Objective function g can be expressed by a kernel function. Since RBF function is used as the kernel function in this paper, there is only one parameter for g.

Some authors select the kernel parameters of SVM only by the inter-cluster distance in the feature space [7]. In this paper, we used this idea of parameters optimization in LS-SVM. Experiments results show that the method is as good as or better than the proposed method.

III. The Novel LS-SVM Parameter Selection Algorithm

Let the training data set be L, Validation data set be M, test data set be N, the two classes be two labeled clusters.

Step 1: For each given kernel parameter, the corresponding g value is calculated by using (11) with the training data set. Then pick the kernel parameter which leads to minimum value of g as the optimal kernel parameter σ^*.

Step 2: For each γ of the given parameter search interval, combining with the optimal kernel parameter σ^* ,we train a classifier with the training data set, and then use the classifier to classify the validation data set, select the optimal parameter

combination $< \sigma^*, \gamma^* >$ which leads to the classifier with highest validation accuracy for testing process usage.

IV. EXPERIMENTAL RESULTS

A. Description of experimental dataset

In this section, we present some experimental results on a suite of ten data sets from UCI [13] benchmark repository. Some of these data sets are multi-class, so we changed it to binary-class by artificial combination.

The specifications of these data sets are listed in Table Ⅰ.

TABLE I. TEN DATASETS

Data set	*Size*	*Dim*	*Classes*
Blood	748	5	2
Hayes	132	5	3
Lung	32	57	3
New	215	6	3
Pima	768	9	2
Soybean	47	36	4
Waveform	5000	22	3
Breast	699	11	2
Glass	214	10	7
Haberman	306	4	2

B. Parameters analysis of LS-SVM

RBF kernel function is used in our experiments, where σ is set with the values $\left[2^{-20}, 2^{-19}, \cdots, 2^{20}\right]$ and γ is given with the values $\left[2^{-7}, 2^{-6}, \cdots, 2^{7}\right]$. The given range of parameters can cover the most of the effective parameter search space which lead to a high validation accuracy [7].

C. Description of experimental Results

Table Ⅱ shows the optimal parameter combination, training time and testing accuracy of the LS-SVM with the grid search method. Table Ⅲ shows the training time and testing accuracy of the LS-SVM with the proposed method. Table Ⅳ shows the optimal parameter combination, training time and testing accuracy of the LS-SVM with the method by the inter-cluster distance. Table Ⅴ shows comparison between the grid search method and the proposed method in training time. Figure1 shows the relationships of test accuracy with different regularization parameters on ten data sets by the proposed method.

D. Experimental Conclusion

For the proposed method, the training time is the sum of the time of calculating the g values for all kernel parameters and training LS-SVM for all γ with the best kernel parameter combination chosen according to the g values. According to

TABLE II. THE RESULT OF THE GRID SEARCH METNOD

Data set	$< \sigma, \gamma >$	*Testing accuracy(%)*	*Training time (s)*
Blood	$<2^{12}, 2^{-5}>$	73.6	294.3483
Hayes	$<2^{3}, 2^{5}>$	73	4.9953
Lung	$<2^{18}, 2^{-7}>$	66.67	22.1749
New	$<2^{2}, 2^{-3}>$	95.909	19.4280
Pima	$<2^{11}, 2^{6}>$	77.013	211.7923
Soybean	$<2^{13}, 2^{-7}>$	60	16.2512
Waveform	$<2^{9}, 2^{3}>$	94	30.5030
Breast	$<2^{8}, 2^{1}>$	95.714	297.7735
Glass	$<2^{0}, 2^{0}>$	70	13.9798
Haberman	$<2^{18}, 2^{-7}>$	72.258	39.9077

TABLE III. THE RESULT OF THE PROPOSED METHOD

Data set	$< \sigma, \gamma >$	*Testing accuracy (%)*	*Trainingtime (s)*
Blood	$<2^{20}, 2^{-7}>$	76	120.1813
Hayes	$<2^{4}, 2>$	77	2.4218
Lung	$<2^{-20}, 2^{-7}>$	66.67	0.8187
New	$<2^{4}, 2^{2}>$	92.273	9.8948
Pima	$<2^{19}, 2^{7}>$	71.299	127.6047
Soybean	$<2^{19}, 2^{-7}>$	72	0.9296
Waveform	$<2^{17}, 2^{1}>$	90.4	14.0249
Breast	$<2^{15}, 2^{2}>$	95	104.7484
Glass	$<2^{1}, 2^{-2}>$	75.333	4.9952
Haberman	$<2^{-19}, 2^{-7}>$	73.226	19.5109

TABLE IV. THE RESULT OF THE INTER-CLUSTER DISTANCE METHOD

Data set	$< \sigma, \gamma >$	*Testing accuracy (%)*	*Training time (s)*
Blood	$<2^{9}, 2^{3}>$	76	119.5422
Hayes	$<2^{4}, 2^{6}>$	79	2.464
Lung	$<2^{2}, 2^{-7}>$	66.67	0.8
New	$<2^{3}, 2^{6}>$	94.545	10.0565
Pima	$<2^{6}, 2^{-2}>$	76.36	133.8627
Soybean	$<2^{1}, 2^{-7}>$	70	0.8733
Waveform	$<2^{3}, 2^{-3}>$	92	14.2312
Breast	$<2^{15}, 2^{7}>$	95.57	103.4564
Glass	$<2^{-1}, 2^{3}>$	74	4.7139
Haberman	$<2^{3}, 2^{-7}>$	74.194	20.2156

TABLE V. TRAINING TIME WITH THE PROPOSED MATHODS AND THE GRID SEARCH

Data set	*Grid search Training time (s)*	*Proposed method Training time (s)*	*Speed-up (times)*
Blood	294.3483	120.1813	2.4493
Hayes	4.9953	2.4218	2.06
Lung	22.1749	0.8187	27.0855
New	19.4280	9.8948	1.9627
Pima	211.7923	127.6047	1.6598
Soybean	16.2512	0.9296	17.4931
Waveform	30.5030	14.0249	2.1749
Breast	297.7735	104.7484	2.8427
Glass	13.9798	4.9952	2.7986
Haberman	39.9077	19.5109	2.0454

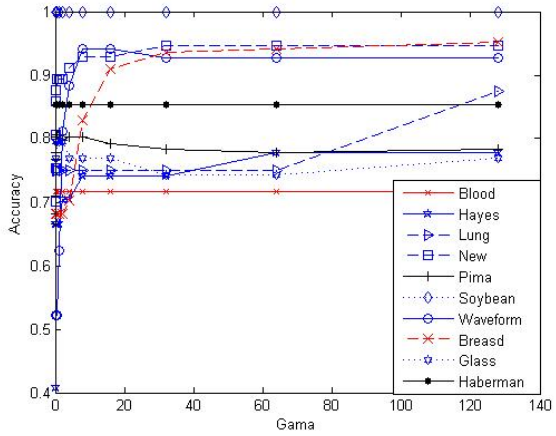

Figure 1. Test accuracy of the LS-SVM model obtained by the proposed method with different regularization parameter γ on ten datasets.

the experiment results, the LS-SVM trained by the proposed method with g performs almost the same as those models chosen by grid search do when using RBF kernels. Mean while, as the kernel parameters are decided by calculating the corresponding g values, not by training all the LS-SVM models, the total training time includes only the computation time of the g values and the LS-SVM training time of deciding a proper γ. Since it is not necessary to train models for the kernel parameters, the training time can be significantly shortened.

V. CONCLUSION

We propose calculating the distance between samples in the feature space to determine the optimal value of the kernel parameter for LS-SVM. Then the optimal regularization parameter is determined by grid search method. Since the proposed method only need to calculate distance for searching the optimal kernel parameter, calculating the distance is a simple and non-iterated process. Thus, the training time of the proposed method to obtain LS-SVM models can be significantly shortened. Since the regularization parameter γ is not incorporate into the proposed strategies, or else the training time might be further minimized. How quickly and effectively select of LS-SVM parameters is still a hot topic.

ACKNOWLEDGMENT

This research is supported in part by the Natural Science Foundation of Hebei Province (No. F2011201063, F2010000323), the National Natural Science Foundation of China (No. 60903088, 60903089), the Scientific Research Project of Department of Education of Hebei Province (No. 2009107).

REFERENCES

[1] V. N. Vapnik, Statistical learning theory. John Wiley, New York, USA 1998.

[2] J.A.K. Suykens, J. Vandewalle. Least squares support vector machine classifiers. Neural Processing Letters(S1370-4621), vol. 9, 1999, pp. 293–300.

[3] J.A.K. Suykens, L. Lukas, J. Vandewalle, Sparse approximation using least squares support vector machines, in: Proceedings of the IEEE InternationalSymposium on Circuits and Systems (ISCAS 2000), vol. 2, 2000, pp. 757−760.

[4] K. Polat, S. Gunes, A novel approach to estimation of E. Coli promoter genes equences: combining feature selection and least square support vector machine (FS_LSSVM), Appl. Math. Comput, vol.190, 2007, pp.1574−1582.

[5] C.-W. Hsu, C.-J. Lin, A comparison of methods for multiclass support vector machines, IEEE Trans.Neural Networks, vol.13,2002,pp.415–425.

[6] S. An., W. Liu, S. Venkatesh, Fast cross-validation algorithms for least squares support vector machine and kernel ridge regression. *Pattern Recognition*, vol.*40*, 2007, pp.2154–2162.

[7] K.-P. Wu, S.-D. Wang, Choosing the kernel parameters for support vector machines by the inter-cluster distance in the feature space, Pattern Recognition: The Journal of the Pattern Recognition Society, vol.42, 2009, pp.710–717.

[8] Avci, E. Selecting of the optimal feature subset and kernel parameters in digital modulation classification by using hybrid genetic algorithm-support vector mahines: HGASVM. Expert System with Applications, vol.36, 2009, pp.1391−1402.

[9] X.-L. Zhang, X.F. Chen, Z.J. He, An ACO-based algorithm for parameter optimization of support vector machines. Expert Systems with Applications. vol.37, 2010, pp.6618−6628.

[10] Y. Bazi, F. Melgani, Semisupervised PSO − SVM regression for biophysical parameter estimation, IEEE Transactions on Geoscience and Remote Sensing, vol.45 ,2007, pp.1887−1895.

[11] T. Van Gestel, J.A.K Suykens, B. Baesens, et al. Benchmarking least squares support vector machine classifiers, Mach. Learn.vol. 54 , 2004, pp.5−32.

[12] D.J. Sebald , J.A. Sebald, Support vector machine techniques for nonlinear equalization. IEEE Transactions On Sit, real Processing, vol. 48, 2000, pp.3 217-3 226.

[13] M. Murphy. UCI-Benchmark Repository lf Artificial and Real Data Set. http: // www. ics. uci. Edu.