Research on Kernel Function of Support Vector Machine

Lijuan Liu[1,2] Bo Shen[1,2*] Xing Wang[3]

[1] School of Electronic and Information Engineering, Beijing Jiaotong University
Beijing, 100044, China
[2] Key Laboratory of Communication and Information Systems, Beijing Municipal Commission of Education, Beijing Jiaotong University
Beijing, 100044, China
{11120126, bshen}@bjtu.edu.cn
[3] China Information Technology Security Evaluation Center
Beijing, China
wangx@itsec.gov.cn

Received 12 May 2013; Revised 23 September 2013; Accepted 20 October2013

Abstract. Support Vector Machine is a kind of algorithm used for classifying data including linear and non-linear, which not only has a solid theoretical foundation, but also is more accurate than other algorithms in many areas of applications, especially in dealing with high-dimensional data. In SVM, kernel function is an important component, which makes it not necessary for us to get the specific mapping function in solving quadratic optimization problem of SVM. The only thing we need to do is to use kernel function to replace the complicated calculation of the dot product of data set, which noticeably reduces the number of dimension calculation. In this paper, we will introduce the theoretical basis of support vector machine, summarize the research status and analyze the research direction and development prospects of kernel function.

Keywords: support vector machine, high-dimension data, kernel function, quadratic optimization

1 Introduction

Support Vector Machine (SVM) was introduced into the field of machine learning and its related area in 1992 [1]. Since then, it receives widespread attention of researchers and has made great progress in many fields. SVM has a solid theoretical foundation and straightforward mathematical model and has got considerable development in pattern recognition, function estimation, time series forecasting and many other areas. It uses a nonlinear mapping to map original training data into high-dimensional space for finding the optimal classification hyper plane that separates those data into different categories. SVM is based on VC dimension (Vapnik Chervonenkis dimension) and structural risk minimization principle of Statistical Learning Theory (SLT) [2, 3]. Compared with traditional neural networks, SVM gains great enhancement in generalization ability and overcomes some problems existing in feed-forward neural networks, such as local minimum and the curse of dimensionality [4]. Further, the introduction of kernel function greatly simplifies the complexity of dot product operation in SVM for nonlinear data classification, and makes it possible to distinguish and enlarge the useful features by SVM. Generalization ability of SVM relies heavily on the choice of kernel function. Basing on kernel function, SVM is playing more and more powerful role in the field of data mining.

2 Support Vector Machine

Suppose D={ $(X_1, y_1),(X_2, y_2),\cdots,(X_{|D|}, y_{|D|})$ } is a linear separable data set，among which X_i is the data with binary class $y_i \in \{-1,+1\}$. For the purpose of data classification, an optimal hyper plane can be denoted as

$$W \bullet X + b = 0 \tag{1}$$

where W is the weight vector of X, and parameter b is a scalar that is often referred to as the bias. The formula $W \bullet X$ stands for the dot product of W and X. In order to make the hyperplane correctly classify input data and have classification interval, it is required to satisfy the following constraint:

$$y_i[(w \bullet x_i + b)] \geq 1 \text{ , } i = 1, 2, ..., l \quad \textbf{(2)}$$

From geometric point of view, the entire input space is divided into two parts by a hyper plane: one with +1 class and another with -1. Obviously, the hyper plane appears a line form when X belongs to a two-dimensional space and a surface form when X is in three-dimensional space. The maximum margin distance between these two parts of data set is $2/\|W\|$. Theoretically speaking, SVM can implement optimal classification for linear separable data.

Thus, the classification task is transformed into finding a hyper plane among all categories. SVM is just such a method to discover the optimum classification hyper plane by means of support vector [5]. Here optimum plane indicates the one that represents maximum margin between classes among all hyper planes that might classify the data set.

For linear separable data like those above, SVM can directly classify them into some categories within the input space. But in the practical application, the vast majority of the sample data set in the original space is not linear, that is to say, linear function cannot be adopted to carry out the correct classification. For nonlinear separable data, SVM has to map the original input data X with nonlinear mapping ($\phi : X \rightarrow F$) into another high-dimensional space where the maximum interval of classification could be solved. This new high-dimensional space is called feature space. In essence, this approach is a method to increase the instruction function set VC dimension.

3 Kernel Function

In recent years, kernel function has received extensive attention, particular due to the increased popularity of the SVM. Kernel function can be served as a bridge from linearity to non-linearity for algorithms that can be served as a bridge from linearity to non-linearity for algorithms that can be expressed in terms of dot product. In feature space, a dot product operation can directly be substituted by kernel function, while we do not have to know the concrete eigenvector and the concrete mapping function, which is also known as kernel trick. Suppose $\phi(x_i) \in R^m$, $m > d$ is the transformation result of $x \in R^d$ where the nonlinear transformation function is $\phi(x)$. Parameter x_i belongs to d dimension input space, and $\phi(x_i)$ belongs to m dimension feature space. By this time, optimization problem is changed into the function like

$$Q(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \alpha_i \alpha_j y_i y_j \{\phi < (x_i), \phi(x_j) >\} \quad \textbf{(3)}$$

Inner product of the formula above is done in a relatively high dimension space, so there may be a dimension disaster problem. After introducing kernel function into SVM, the optimization problem becomes

$$Q(a) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad \textbf{(4)}$$

The corresponding classification function is

$$f(x) = \text{sgn}\{(w \bullet x) + b\} = \text{sgn}\{\sum_{i=1}^{n} \alpha_i^* y_i K(x_i, x_j) + b^*\} \quad \textbf{(5)}$$

Here are several kernel functions.

Linear kernel function: $K(X_i, X_j) = X_i \bullet X_j$. **(6)**

Polynomial kernel function: $K(X_i, X_j) = (X_i \bullet X_j + 1)^h$. **(7)**

Gaussian radial basis function kernel function: $K(X_i, X_j) = \ell^{-\|X_i - X_j\|^2 / 2\sigma^2}$. **(8)**

Sigmoid kernel function: $K(X_i, X_j) = \tanh(\kappa X_i \bullet X_j - \delta)$. **(9)**

Linear kernel function is the most simple kernel function given by the inner product and is often equivalent to non-kernel counterparts. Polynomial kernel function is well suited for problems where all the training data is normalized. In Gaussian radial basis function kernel function, parameter σ plays a major role in the performance of this kernel function, and should be carefully tuned according to the problem at hand. SVM with Sigmoid kernel function is equivalent to a simple 2-layer neural network named multilayer perception without hidden layer. In SVM, node number of hidden layer and the weight of hidden layer to the input node are automatically determined in design process. Besides, SVM can finally obtain the global optimal value, which has quite good generalization ability for unknown sample data.

The performance of SVM mainly depends on model selection, including the selection of the kernel function type and kernel parameter [6]. In the study of kernel function selection, kernel alignment [7] based on the hypothesis that the kernel matrix of a good kernel function should be as similar as possible to the calibration matrix is a good method. Under normal circumstances, the first consideration of choosing kernel function is the Gaussian radial basis function kernel function (RBF) because it has fewer parameters to select and has similar performance to Sigmoid kernel function for some parameters.

The development and application of SVM has been greatly promoted since the introduction of kernel function, and its application area has extended from hand-written numeral recognition, reference time series prediction test and other traditional application area to new areas such as information image processing [8], industrial process control, etc. The following content will center on the discussion of kernel function in SVM and put forward future research directions.

3.1 Kernel Clustering

Clustering analysis divides data objects into different subsets. Data objects in the same subset are similar to each other, and those located in different subsets have different properties. Kernel clustering combines kernel function and clustering together in line with clustering classification characteristics [9]. Kernel clustering clusters training data and test data separately many times, and then constructs the kernel function on the basis of those clustering results. Chapell et al. [10] used different conversion functions to change the eigenvalue decomposed by kernel matrix in order to obtain diverse kernel functions, thus proposed an overall framework of constructing kernel clustering. In the light of big time complexity problem in kernel clustering raised by Chapell, Jason Weston et al. came up with the bagged clustering kernel [10]. Although the bagged clustering kernel shortened kernel clustering time, there is still much room for improvement in terms of classification accuracy.

Suppose that the feature space of input data after mapping is $\phi(x_1)$, $\phi(x_2)$, …, $\phi(x_l)$. Euclidean distance of feature space is

$$d_H(x, y) = \sqrt{\|\phi(x) - \phi(y)\|^2} = \sqrt{\phi(x) \bullet \phi(x) - 2\phi(x) \bullet \phi(y) + \phi(y) \bullet \phi(y)} \quad \textbf{(10)}$$

The kernel function into the above formula, we can obtain the Euclidean distance of feature space as:

$$d_H(x, y) = \sqrt{K(x, x) - 2K(x, y) + K(y, y)} \quad \textbf{(11)}$$

Kernel clustering can adopt the fuzzy clustering partition method and hard division. Fuzzy C-means (FCM) clustering algorithm [11] introduces fuzzy set theory to the process of clustering. Fuzzy kernel clustering algorithm firstly maps the input data into high-dimensional feature space to enlarge pattern differences between categories, and then carry through fuzzy clustering in feature space [12, 13]. Fuzzy kernel clustering algorithm is able to highlight differences among different sample characteristics and increase clustering accuracy and speed [14]. Appropriate kernel function and parameters can be found in different situations such as different clustering prototypes, different density and high dimensional data. Selecting appropriate kernel function parameters and reducing time complexity problem remains to be solved.

The objective function of fuzzy clustering algorithm in feature space is expressed as

$$J_m(U,v) = J_m(U,\beta_1,\beta_2,...,\beta_c) = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^m \left\| \phi(x_k) - \sum_{l=1}^{n} \beta_{il\phi}(x_l) \right\|^2 \quad \mathbf{(12)}$$

$$\left\| \phi(x_k) - \sum_{l=1}^{n} \beta_{il}\phi(x_l) \right\|^2 = \phi(x_k)^T\phi(x_k) - 2\sum_{l=1}^{n}\beta_{il}\phi(x_k)^T\phi(x_l) + \sum_{l=1}^{n}\sum_{i=1}^{n}\beta_{il}\phi(x_k)^T\beta_{ij}\phi(x_j) \quad \mathbf{(13)}$$

Bring the kernel function to the above formula, we can obtain

$$J_m(U,\beta_1,\beta_2,...,\beta_c) = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^m (K_{kk} - 2\beta_i^T K_k + \beta_i^T K \beta_i) \quad \mathbf{(14)}$$

Where parameter c is clustering number initially set; and v_i is initialized clustering center; and u_{ij} is membership function of sample j belonging to category i . $U = \{u_{ij}\}$, $v = \{v_1, v_2, ..., v_c\}$, and $n > 1$ is weighted index. The criterion of fuzzy clustering algorithm is the minimum value of the above objective function until each membership value stabilized. The algorithm, similar to the SVM classification hyperplane solution, uses dual representation of the cluster centers to obtain clustering results in feature space, which is a natural extension of FCM. Input space fuzzy kernel clustering algorithm objective function is

$$J_m(U,v) = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^m \left\| K(x_k,x_k) + K(v_i,v_j) - 2K(x_k,v_i) \right\| \quad \mathbf{(15)}$$

Thus the execution of FCM in Euclidean distance is expanded to different distance measurements in the same space of new clustering.

Multi-class classification problem often occurs in reality, while SVM is proposed for the binary classification problem. So it is an important research problem to expand SVM classifier to multi-class classification field [15, 16]. There are roughly two types in SVM multi-class classification methods: one is disposable algorithm and the other one is decomposition and reconstruction algorithm. Disposable algorithm is used to solve the large quadratic programming problem in all the training samples, and at the same time separate all categories apart. This method has multiple variables and high computational complexity and cannot obtain satisfactory classifying effect on training speed and classification accuracy when there are multiple categories. Decomposition and reconstruction algorithm converts multi-class classification problem into a binary classification problem. Zhao et al. [17] apply fuzzy kernel clustering to multi-class classification method and use a tree structure to combine multiple SVM classifier together to achieve multi-class classification. This method firstly uses fuzzy kernel clustering to generate fuzzy class to dig on the periphery of the fuzzy class and information between different fuzzy classes, and then reduces the scale of training samples. Multi-class classifier effectively solves the serious overlap problem existing in traditional fuzzy classes.

Kernel clustering support vector machine has been successfully utilized in biomedical, text classification [18] and many other applications, and undoubtedly will involve a wider range in the future.

3.2 Super-Kernel Function

SVM for data classification often brings about two dilemmas using one single kernel function: one is unable to complete effective nonlinear mapping; the other one is over-fitting or under-fitting [19]. The extrapolation ability of Gaussian radial basis function kernel function weakens along with the increase of σ [20], so it has strong locality. Polynomial kernel function regulates different mapping dimensions through adjusting h parameter, and calculation quantity grows with h parameter, thus having strong global property and poor locality.

In order to adapt to the increase of data set and high efficiency requirements, many algorithms have been developed, such as large data set training [21] algorithm, super kernel learning [22] algorithm and fast convergence algorithm [23]. According to the Mercer theorem, nonnegative linear combination of Mercer kernel and the product of Mercer kernel and super-kernel function are still Mercer kernels. Combining several different kernel functions with a polynomial composition can give full play to the excellent characteristics of kernel functions in dealing with data classification, and overcome disadvantages of one single kernel function. Super-kernel function can maintain translational invariance and rotation invariance, which provides a fresh effective way to the study of kernel function construction in SVM. A simple form of super-kernel function is like the following formula:

$$K(X_i, X_j) = \beta_1 \ell^{-\|X_i - X_j\|^2 / 2\sigma^2} + \beta_2 (X_i \bullet X_j + 1)^h \tag{16}$$

It is the linear combination of Gaussian radial basis function kernel function and h polynomial kernel function. The commonly used kernel function performance comparison is shown in the following figure.

Table 1. The commonly used kernel function performance comparison

Kernel functions	Range of application	Sensibility to parameters
Linear kernel	Overall situation	Insensitivity
Polynomial kernel	Overall situation	Non-stationary
Gaussian kernel	locality	Sensitive to σ
Exponential kernel	locality	Sensitive to σ
Sigmoid kernel	Overall situation	Sensitive to data dimension
Multiquadric kernel	locality	stationary
Circular kernel	Overall situation	stationary

We can have a much better classification effect on data set if we combine different kernel functions to build one super-kernel function reasonably according to their application range and characteristic of sensitivity to parameters. Super-kernel function parameters were regulated in the form of parameter vector, that is to say, super-kernel function adjusts all parameters at the same time, and those several more parameters just add to the length of parameter vector but do not affect the determination time of parameters.

3.3 Kernel Parameter Selection

According to different kernel selection methods, it can be divided into construction of kernel and kernel parameter adjustment problem. Kernel construction means constructing more useful kernel and make use of kernel function and the nature of kernel matrix by means of calculation. The most commonly used kernel function such as polynomial kernel function and Gaussian radial basis function kernel function are also constructed by calculations. Selection of kernel parameter has direct impacts on the performance of SVM classifier. A commonly used parameter selection method is based on the Generalization Error estimates [24]. Generalization Error estimates predicate and forecast generalization capability of classification decision criteria by means of training data set [25]. Vapnik et al. [26] estimate generalization ability by span of support vectors on the basis of Generalization Error estimates, which is of higher accuracy but has more complex calculation process. Men et al. [27] put forward a method of using data set to estimate the optimal kernel parameters. This method starts from the point of geometry to calculate the appropriate convex hull of sample data and then selects the appropriate kernel parameters by directly calculating the minimization ratio between the smallest class of ultra-sphere radius and the maximum interval of all samples.

Kernel principal component analysis (KPCA) is a nonlinear method of converting raw data into feature space and feature extraction is the key to selecting optimal nonlinear transformation kernel function parameters. How to choose the optimal or near optimal nonlinear transform of kernel function parameter and make the separable measure largest is the key of KPCA to apply to the area of feature extraction. At present, it is still an unsolved problem to choose the optimal kernel function and parameters according to specific problems to achieve the optimal classification effect. The main technical approaches is to map the input data set x_k to $\phi(x_k)$ of the feature space by means of nonlinear mapping function $\phi : R^N \to F$, and then execute PCA in feature space.

Qi et al. [28] proposed a kernel parameter selection method to solve LOO (leave-one-out) upper bound minimum point based on genetic algorithm, where we can choose the reproduction operator and combine genetic algorithm with steepest descent method, improving forecast accuracy but not leading to local optimal solution. Chen et al. [29, 30] adopt different generalization ability estimates as the fitness function of genetic algorithm. This method not only reduces calculation time in choosing parameters but also reduces dependence on initial values.

Kernel matrix measures similarity between kernels by means of kernel alignment to obtain the distribution situation of samples in feature space, thus acquiring useful information of input vector in feature space with feature mapping function unknown. Parameter selection method based on kernel matrix similarity measure starts from research on kernel matrix in order to search for optimal kernel parameters and learning models, and this method can improve calculation speed of SVM. Liu Xiangdong et al. [31] eventually found the optimal kernel parameters and kernel matrix by means of experimenting on UCI standard data set and FERET standard faces library. Parameter selection method based on kernel matrix similarity measure can serve as a feasible method to choose optimal SVM model, and it also has definite reference value for choosing other kernel parameters.

4 Conclusion

The study of kernel function is an important data mining research content, so choosing appropriate kernel function and parameters can give full play to the performance of SVM and even has remarkable significance in promoting the popularization and application of data mining. This paper does research on kernel function of SVM and makes summary comments on kernel clustering, super kernel function and selection of kernel parameters. Judging from current study, the author believes that the study of kernel function in the following areas is to be further developed:

(1) Realizing data mapping efficiently and reliably in the environment of big data. "Big data" has features of magnanimity, high growth rate and diversification, and requires new processing mode to excavate useful information from the massive data and get an insight in them. In this case, the conventional kernel function transformation will face with new bottleneck in processing speed and quality. On the basis of existing research, it remains further study to extend the scale of kernel function processing data and to select appropriate kernel parameters and to further improve processing quality and speed.

(2) Giving full play to the advantages of different kernel functions in super-kernel functions. On the perspective of present research on selection of kernel parameters, Gaussian radial basis function (RBF) and its parameter selection have been detailed studied in the field by virtue of its favorable advantages in computer vision. How to expand the application area of polynomial kernel function and Sigmoid kernel function, especially give full play to the advantage of each kernel function still needs deeper exploration. It is worth studying to select and optimize super-kernel function parameters and apply concept of constructing super-kernel function to SVM.

(3) Selecting appropriate kernel function of SVM for specific application field. The range of data is developing from structured data to the direction of semi-structured and unstructured. As one part of data mining classification algorithm, application fields of SVM continue to expand, thus it appears particularly important to select appropriate kernel functions of specific domain. The choice of kernel function is closely related to data field [32], and in the meantime, the performance of kernel function depends largely upon the selection of parameters. Future studies are required to determine appropriate kernel function and its parameters according to different application areas of SVM for the sake of reducing storage space consuming and computing time of computers.

Big data processing is the future research tendency, and with the arrival of cloud era, big data is attracting more and more attention. In future work, the author will mainly focus on the study of achieving efficient and accurate classification with SVM under the environment of big data.

Acknowledgement

This work has been supported by the National Natural Science Foundation of China under Grant 61172072, 61271308, and Beijing Natural Science Foundation under Grant 4112045, and the Research Fund for the Doctoral Program of Higher Education of China under Grant W11C100030, the Beijing Science and Technology Program under Grant Z121100000312024.

References

[1] B. Boser, I. Guyon, V. Vapnik, “A training algorithm for optimal margin classifiers,” in *Proceedings of the 5th Annual ACM Conference on Computational Learning Theory*, pp.144-152, 1992.

[2] V. N. Vapnik, *The Nature of Statistical Learning Theory*, translated by Zhang Xuegong, Beijing: Tsinghua University Press, 2000.

[3] V. N. Vapnik, *Statistical Learning Theory*, translated by Xu Jianhua, Zhang Xuegong, Beijing: Publishing House of Electronics Industry, 2004.

[4] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer Press, 1995.

[5] C. Ju, F. Guo, "A distributed data mining model based on support vector machines DSVM," pp. 1855 – 1863, 2010.

[6] S. Zhu, R. Zhang, "Research for selection of kernel function used in support vector machine," *Science Technology and Engineer*, Vol. 8, No. 16, pp. 4513-4517, 2008.

[7] N. Cristianini, J. Kandola, et al., "On kernel target alignment," in *Proceedings of Neural Information Processing Systems*, MIT Press, Cambridge, MA, pp. 367-373, 2002.

[8] Z. Yang, "Kernel-based support vector machines," *Computer Engineer and Applications*, Vol. 44, No. 33, pp.1-6, 2008.

[9] T. Li, X. Wang, "A semi-supervised support vector machine classification method based on cluster kernel," *Application Research of Computers*, Vol. 30, No. 1, pp. 42-45, 2013.

[10] C. Tison, J. M. Nicolas, F. Tupin, et al, "A new statistical model for Markovian classification of urban areas in high-resolution SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 42, No. 10, pp. 2046-2057, 2004.

[11] J. C. Bezdek, *Pattern Recognition with Fuzzy objective Function Algorithms*, New York: Plenum Press, 1981.

[12] Z. Wu, X. Gao, W. Xie, "A study of a new fuzzy clustering algorithm based on the kernel method," *Journal of Xi 'an University of Electronic Science and Technology Magazine*, Vol. 31, No. 4, pp. 533-537, 2004.

[13] N. Zhang, Y. Zhang, "Support vector machine ensemble model based on KFCM and its application," *Journal of Computer Applications*, Vol. 30, No, 1, pp. 175-177, 2010.

[14] W. Cao, Y. Zhao, S. Gao, "Multi-class support vector machine based on fuzzy kernel clustering," *CIESC Journal*, Vol. 61, No. 2, pp. 420-424, 2010.

[15] C. Angulo, X. Parra, A. Catala, "K-SVCR. A support vector machine for multi-class classification," *Neurocomputing*, Vol. 55, No. 9, pp. 55-77, 2003.

[16] J. C. Platt, N. Cristianini, J. Shawe-Taylor, "Large margin DAGs for multiclass classification," *Advances in Neural Information Processing Systems*, Vol. 12, No. 3, pp. 547-553, 2000.

[17] H. Zhao, L. Rong, "SVM multi-class classification based on fuzzy kernel clustering," *Systems Engineering and Electronics*, Vol. 28, No. 5, pp. 770-774, 2006.

[18] Z. Yang, "Research Progress of the kernel function support vector machine," *Science & Technology Information*, Vol. 19, pp. 209-210, 2008.

[19] L. Jia, S. Liao, "Support vector machines with hyper-kernel functions," *Computer Science*, Vol. 35, No.12, pp. 148-150, 2008.

[20] L. Guo, S. Sun, X. Duan, "Research for support vector machine and kernel function," *Science Technology and Engineering*, Vol. 8, No. 2, pp. 487-489, 2008.

[21] R. Collobert, S. Bengio, "SVMTorch: Support vector machines for large-scale regression problems," *Journal of Machine Learning Research*, Vol. 1, pp. 143-160, 2001.

[22] S. O. Cheng, A. J. Smola, R. C. Williamson, "Learning the kernel with hyper-kernel," *Journal of Machine Learning Research*, Vol. 6, pp. 1043-1071, 2005.

[23] J. Platt, C. J. C. Burges, "Fast training of support vector machines sequential minimal optimization," *Sholkpof B, Smola A J, eds*, MIT Press, 1998.

[24] T. Wu, *Kernels' Properties, Tricks and Its Application on Obstacle Detection*, Changsha: National University of Defense Technology, 2003.

[25] Y. Fu, D. Ren, “Kernel function and its parameters selection of support vector machines,” *Science and Technology Innovation Herald*, Vol. 9, pp. 6-7, 2010.

[26] O. Chapelle, V. Vapnik, O. Bousquet, et al, “Choosing multiple parameters for support vector machines,” *Machine Learning*, Vol. 46, No. 1, pp. 131-159, 2002.

[27] C. Men, W. Wang, “Kernel parameter selection method based on estimation of convex,” *Computer Engineer and Design*, Vol. 27, No. 11, pp. 1961-1963, 2006.

[28] Z. Qi, Y. Tian, Z. Xu, “Kernel-parameter selection problem in support vector machine,” *Control Engineering of China*, Vol. 12, No. 44, pp. 379-381,2005.

[29] P. W. Chen, J. Y. Wang, H. M. Lee, “Mode selection of SVNs using GA approach,” in *Proceedings of 2004 IEEE International Joint Conference on Neural Networks*, IEEE Press, pp. 2035-2040, 2004.

[30] C. H. Zheng, et al., “Automatic parameters selection for SVM based on GA,” in *Proceedings of the 5th World Congress on Intelligent Control and Automation*, IEEE Press, pp. 1869-1872, 2004.

[31] X. Liu, B. Luo, Z. Qian, “Optimal model selection for support vector machines,” *Journal of Computer Research and development*, Vol. 42, No. 4, pp. 576-581, 2005.

[32] T. Wang, J. Chen, “Survey of research on kernel selection,” *Computer Engineer and Design*, Vol. 33, No. 3, pp. 1181-1186, 2012.