

<http://ansinet.com/itj>

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSI*net*

Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Flood Disaster Evaluation Model Based on Kernel Dual Optimization Support Vector Machine

^{1,2}Weiping Deng, ¹Jianzhong Zhou, ¹Qiang Zou, ¹Jian Xiao, ¹Yongchuan Zhang and ³Weihua Hua

¹School of Hydropower and Information Engineering,

Huazhong University of Science and Technology, Wuhan, 430074, China

²School of Information and Management, Hubei University of Economics, Wuhan, 430205, China

³Faculty of Information Engineering, China University of Geosciences, Wuhan, Hubei, 430074, China

Abstract: Support vector machine is adopted in this paper to construct flood disaster evaluation model, which can be indicated as a comprehensively nonlinear classification issue. In this article, kernel function of SVM is optimized by kernel transformation and kernel parameters optimization. In order to discriminate the flood disaster evaluation indexes and really reflect their classification contributions, kernel function is weighted to promote classification performance and reduce the error influence by weak features. Further more, after analyzing the over learning issue of traditional grid search, an improved grid search is proposed to optimize the kernel parameters. The new search method illustrates it is more reasonable to make second search around some suboptimal solutions. By this search method, the final solution obtains high accuracy in testing samples and reduces the value of penalty factor. The experiment results show the promotion of classification precision by the dual optimization model and identify it could be a good choice for other classification issues.

Key words: Support vector machine, kernel optimization, grid search, feature weighted, flood disaster evaluation

INTRODUCTION

Support vector machine (SVM) is a new learning algorithm based on Statistical Learning Theory. It keeps to structural risk minimization principle and overcomes shortcoming of empirical risk (Vapnik, 2000). It does well in solving nonlinear problem by mapping the sample from low dimensional space to high dimensional space. Besides, it is widely adopted in many classification and regression problems by its excellent generational ability. Flood disaster grade evaluation is a comprehensive classification issue (Jiang *et al.*, 2009) so SVM is adopted to evaluate disaster loss grade in this paper.

Kernel function is the key structure of SVM because it indirectly reflects the special distribution of samples in high dimensional space (Cristianini and Shawe-Taylor, 2000). Then samples' special relations will directly influence the classification result. So how to select and optimize the kernel function is significant for SVM.

The general optimization method for kernel function is to search for optimal kernel parameters. Obviously, selecting better kernel parameters will gain better performance. Many methods have been adopted to search for optimal kernel parameters, such as grid search method (Wang *et al.*, 2007, 2012; Kapp *et al.*, 2012), Particle Swarm Optimization (PSO) (Lin *et al.*, 2008; Ren and

Huo, 2009; Jiang *et al.*, 2010), Genetic Algorithm (GA) (Adankon and Cheriet, 2010; Chen, 2011), Ant Colony Optimization (ACO) (Jiang and Liu, 2012), Chaos algorithm (Yuxia and Hongtao, 2012) etc. Most of intelligent algorithms are complex and easy to fall into local optimum. Grid search method costs lots of time and its precision is not high. So this paper modify grid search to increase accuracy and save time.

Optimizing kernel parameters is significant, further more; people find the efficient way for optimizing the kernel function is to design it anew according to the practical problem. Such as flood disaster loss assessment, the importance degree of each assessment index is markedly different so their contributions to classification should be discriminated. But traditional SVM treat all feature indexes the same and ignore the diversity of indexes. Some researches try to solve this problem, feature weighted SVM (Wang *et al.*, 2009) is proposed to reduce the classification influence of weak feature; literature (Shi and Yang, 2008) disposes classification and recognition by feature selection. They are efficient way to solve practical issue but the kernel parameters have not been optimized.

According to the two aspects of optimization, this paper proposes the feature weighted SVM based on improved grid search. Firstly, calculate reasonable feature

weights according to practical issue. Then modify the kernel function by weighting all feature indexes. Lastly, optimize the kernel parameter by the improved grid search. This method does not only develop the kernel function itself but also propose an efficient way to search for the optimal kernel parameters.

THEORY OF SUPPORT VECTOR MACHINE

SVM method is established on the structural risk minimization principle of statistical theory (Vapnik, 2000). It aims to find a computational way to express optimal separating hyper-plane in a high dimensional feature space which satisfied all the constraints. Suppose there is data set $\{(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)\}$, $x_i \in R^m$, $y_i \in \{-1, 1\}$. N represents the number of data samples and m represents the dimension of sample space. Whether y is +1 or -1 represent two classes of different samples. Let $\langle w, x \rangle$ denotes the inner product of the vectors w and x . Based on theory of minimum risk, SVM aims to find out a function to construct the classification hyper-plane denoted by:

$$\langle w, x \rangle + b = 0 \tag{1}$$

The hyper-plane needs not only to separate the samples without error but also to make the separation distance maximum. The hyper-plane which has maximal distance between two sides of support vector points is the optimal classification hyper-plane. Besides, the two class margin planes which constructed by support vector points can be expressed as $\langle w, x \rangle + b = \pm 1$ and the data samples in the sample space can be expressed as:

$$y (\langle w, x \rangle + b) \geq 1 \tag{2}$$

w represents the normal vector and b represents the excursion of hyper-plane. So the distance between support vector points and optimal hyper-plane is $\frac{1}{\|w\|}$. The interval between two sides of margin hyper-plane is $\frac{2}{\|w\|}$.

Obviously, to maximize the distance is equal to minimize $\|w\|$. Then how to find the optimal classification high-plane is transferred to solve the quadratic programming problem denoted by:

$$\min(w) = \frac{\|w\|^2}{2} \tag{3}$$

Subject to $y_i (\langle w, x_i \rangle + b) \geq 1, i = 1, 2, \dots, n$.

By the conversion of LaGrange factor, the above quadratic programming problem is transformed into the following dual problem which is formulated as:

$$\max Q(a) = \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n a_i a_j y_i y_j \langle x_i, x_j \rangle \tag{4}$$

With constraints:

$$\sum_{i=1}^n a_i y_i = 0$$

and $a_i \geq 0, i = 1, 2, \dots, n$.

By solving the optimal solution, the discriminant function of optimal hyper-plane is presented as:

$$f(x) = \text{sgn} \left\{ \sum_{i=1}^n a_i^* y_i \langle x_i, x \rangle + b \right\} \tag{5}$$

In this formula, a_i^* is the optimal LaGrange multiplier, x_i is the corresponding support vector.

Furthermore, in order to solve the non-linearly separable problem, the slack variables are introduced in the classifier. The Eq. 3 is rewritten by:

$$\min(w, \xi) = \frac{\|w\|^2}{2} + c \sum_{i=1}^n \xi_i \tag{6}$$

Subject to $y_i (\langle w, x_i \rangle + b) \geq 1 - \xi_i, i = 1, 2, \dots, n$. where $\xi_i, i = 1, \dots, n$ are slack variables and c is the penalty parameter to control the punishment degree on error.

In general, most of the problems are linearly inseparable so the mapping function $\phi(x)$ is applied to map data from low dimensional input space to high dimensional space where the data are sparse and possibly more separable. Then Eq. (5) can be replaced by Eq. 7:

$$f(x) = \text{sgn} \left\{ \sum_{i=1}^n a_i^* y_i \langle \phi(x_i), \phi(x) \rangle + b \right\} \tag{7}$$

According to Mercer condition, a kernel function $K(x_i, x)$ can be found to replace the inner product $\langle \phi(x_i), \phi(x) \rangle$. So the final discriminant function of SVM can be expressed as:

$$f(x) = \text{sgn} \left\{ \sum_{i=1}^n a_i^* y_i K(x_i, x) + b \right\} \tag{8}$$

where, $K(x_i, x)$ is the kernel function which replace inner product in high dimensional space with functional operation in low dimensional space. This paper employs the Gaussian kernel function to solve classification problem, denoted as:

$$K(x, x_i) = \exp(-g \|x - x_i\|^2) \tag{9}$$

In this function g is related to the kernel width which influences the kernel performance directly.

WEIGHTING KERNEL FUNCTION

Kernel function is the key construction of SVM because it indirectly describes the inner product operation in high dimensional space and reflects the spatial relations between samples by Euclidean distance. But the inner product operation is calculated by all the feature values of samples. Some features play main role in the classification while some features play minor role. So it is unreasonable to treat all features equivalent in the traditional SVM (Wang *et al.*, 2011).

Therefore this paper improves the kernel function by weighting feature to reflect the different roles of feature indexes. The main feature will be endowed with more weight while the minor feature will be endowed with lesser weight. Because kernel function reflects the inner product and the Euclidean distance between samples, weighting kernel function is equal to weight the Euclidean distance.

Suppose $X = \{x_i | i=1, 2, \dots, n\}$ is the sample set mapped into the feature space and x_{ik} express the k-th feature value of the i-th sample. $W = (w_1, \dots, w_k, \dots, w_n)$ is the weight vector, w_k express the k-th feature's weight. Then the weighted Euclidean distance between samples x_i and x_j can be formulated as:

$$D_w(x_i, x_j) = \sum_{k=1}^n w_k \|x_{ik} - x_{jk}\| = \sum_{k=1}^n \sqrt{w_k (x_{ik} - x_{jk})^2} \quad (10)$$

If w_k is greater, the vector distance between x_{ik} and x_{jk} will be stretched and the spatial vector will increase the migration on the specific feature, the more classification contribution will be supplied by the feature. Otherwise if w_k is smaller, the spatial vector will reduce the migration on the specific feature and the classification contribution supplied by the feature will be also weaker. Obviously, the weighted Euclidean distance will change the samples' spatial distribution and may also change sample's classification relations. It is essential to revise kernel function by feature's weight. Then define $K(x)$ is the kernel function based on data sets $X \times X$, w is the feature weighted matrix and n is the dimensionality of input space. The weighted kernel function is defined as follows:

$$K_w(x_i, x_j) = K(x_i^T w, x_j^T w) \quad (11)$$

The weighted RBF kernel function can be defined as:

$$K_w(x_i, x_j) = \exp\left(-\frac{\|x_i^T w - x_j^T w\|^2}{2}\right) = \sum_{k=1}^n \sqrt{w_k (x_{ik} - x_{jk})^2} \quad (12)$$

The weighted kernel function can reflect the samples' spatial distribution objectively and highlight the main

features, fade the minor features. This will help to construct the classification hyper plane more reasonably and improve the classification performance.

SVM KERNEL PARAMETERS OPTIMIZATION

In SVM, It is necessary to determine and optimize the kernel parameters to improve performance. In Gaussian kernel function, it relates to penalty factor c and the kernel parameter g . Penalty factor c shows the penal degree when face the classification error. When c is greater, the punishment is stronger and the future points' influence on support vector is more remarkable. So it will lead to over learning state by noise or outliers when c is too great, which make classification not ideal. Parameter g indicates the breadth of the kernel function in the feature space, which directly influences the sample's spatial location. Therefore finding the optimal parameters is the learning and optimization course of SVM. In this paper, improved grid method and the cross-validation method are combined to optimize kernel function.

k-Cross validation: The training data set is divided into k parts of equal subsets by k-Cross Validation method (Zhang *et al.*, 2008). At every turn selecting one part of data as testing data while the other $k-1$ parts of data as training data, repeat the validation course k times to ensure every subset is tested. Calculate the average of the classification accuracy after k times of validation can obtain the single accuracy.

Improved grid search method: The values of C and g are respectively delimited in different ranges as $c = [c_1, c_2]$; $g = [g_1, g_2]$. Divide the two intervals into n equal parts then to form a two dimensional grid, which reduce the interval range for each search. Through ergodic search each pair of c and g can find out the optimal solution.

In practice, people always subdivide the grid to search for better accuracy. But it is identified by many practices that merely narrowing step length and enhancing the density of the grid search can not significantly improve the search accuracy of grid method. On the contrary, this method leads to huge amount of calculation, costing lots of time while the accuracy is not high. Similarly, if only subdivide the grid around the optimal solution which is supported by first search, searching precision doesn't necessarily improve obviously. This is because the optimal solution is always the result of over learning, which maintain the classification precision by increasing the penalty factor c . this will lead the parameter model succeed in training model but fail in test set. After repeated experiments, it is

found out that there are some suboptimal solutions in the first search; Their classifications are similar with the optimal solution while their penalty factors are not high. If set reasonable search range around the suboptimal solution and subdivide the search range can find out better solution.

So this paper abandon the two common search method, improve the search style anew. Steps of the new searching method are as follows:

- Step 1:** Collate sample data and normalize them
- Step 2:** Divide the data set into training set and testing set
- Step 3:** Initialize the parameters' search ranges and look for the first optimal parameter by grid method in the training set. Cross-validation method is used to calculate the average classification accuracy of each pair of parameter. The highest classification accuracy solution is the optimal solution. At the same time, find out several suboptimal solutions whose classification accuracies are close to the optimal solution's accuracy
- Step 4:** Respectively set the reasonable neighborhood ranges of first optimal solution and the several suboptimal solutions. Then reduce the step length and subdivide the grid to search secondly in these new ranges. Cross-validation method is applied again to calculate classification accuracy. Soon find out every optimal solution of each group
- Step 5:** Train the SVM by every optimal parameter solution of each group, which are supplied by step 4. Respectively calculate the classification accuracy of training set by each optimal solution
- Step 6:** Select the highest accuracy solution as the final solution; if solutions' accuracy are same, select the solution whose penalty factor c is smallest as the final solution. Then the optimal parameter group of SVM is determined
- Step 7:** Predict the test set by optimal parameters model to test the performance of model

By this method, for each pair of grid search N-fold cross-validation is used to reasonably divide the training sets and make the train equalization. Grid method can search all nodes while cross-validation method can make training sample equilibrium. Besides, the improved grid method changes the searching tradition, through searching the range of suboptimal solutions to find real optimal solution. This new method plays the advantages of both methods and improves the search accuracy.

FLOOD DISASTER EVALUATION BASED ON SVM

This paper chooses the historical flood disaster data from 1950 to 2009 as samples, chooses the assessment results from paper (Li *et al.*, 2012) as reference values. The flood disasters are classified into four levels: small, medium, large and extreme. The feature assessment indexes are disaster area, inundated area, dead population, and collapsed houses. This paper selects LIBSVM toolbox (Chang and Lin, 2007) as the SVM development tool, applies the improved SVM model and randomly selects 30 samples as training set, other 30 samples as testing set.

Weighting the kernel function: How to ascertain the indexes weights is essential to revise the kernel function. According to the specialty of flood disaster, the importance degree of each assessment index is certain and it will not change by different samples. So the paper adopts Analytic Hierarchy Process to calculate subjective weight of four features. Define the four features' importance ratio as: 1:2:9:3; then construct the pairwise judgment matrix as:

$$M = \begin{pmatrix} 1 & 1/2 & 1/9 & 1/3 \\ 2 & 1 & 1/5 & 1/2 \\ 9 & 5 & 1 & 3 \\ 3 & 2 & 1/3 & 1 \end{pmatrix}$$

The maximum Eigen value λ_{max} is calculated as 4.0080 and the features' subjective weight is (0.065506, 0.119013, 0.604258, 0.211223). In order to ensure the correctness of the judgment matrix, it is necessary to do the consistency checking. The consistency index $CI = (4.0080 - 4) / (4 - 1) = 0.002667$ and the average random consistency index is looked up as 0.9 in Table 2. So the consistency proportion can be calculated as $CR = CI / RI = 0.00267 / 0.9 = 0.002963 < 0.1$. The consistency proportion is much less than 0.1 so the consistency of judgment matrix and the weight can be acceptable and reasonable. Then revise the RBF kernel by weight.

Optimization of kernel parameter: It is pivotal to search for optimal kernel parameters after revising the kernel function. Firstly, the traditional grid search method is adopted and the search result is shown as Fig. 1 and Table 1.

C is the penalty factor of SVM; g reflects the width of kernel function. Suppose the search range of c and g are both $[2^{-10}, 2^{+10}]$ and step length is 1. After first searching, the optional parameter solution is ($c = 1024, g = 2$). The SVM will be over learning state if c is too large and the classification accuracy of testing set will also be low. In

Table 1: Result based on traditional grid search method

Method	Search range of C	Search range of g	Step length	Optimal solution	Accuracy of training set (%)	Accuracy of testing set (%)
Basic grid method	$[2^{-10}, 2^{10}]$	$[2^{-10}, 2^{+10}]$	1.00	C = 1024, g = 2	88.3333	83.3333
Subdivision grid method(1)	$[2^0, 2^{10}]$	$[2^{-5}, 2^{+10}]$	0.10	C = 891.444, g = 1.741	90	86.6667
Subdivision grid method(2)	$[2^0, 2^{10}]$	$[2^{-5}, 2^{+10}]$	0.01	C = 515.561 g = 1.4142	88.3333	83.3333
Subdivision grid method based on optimal solution	$[2^0, 2^{10}]$	$[2^0, 2^1]$	0.01	C = 515.561 g = 1.4142	88.3333	83.3333

Table 2: Result based on regional searching on suboptimal solutions

Method	Search range of C	Search range of g	Step length	Optimal solution	Accuracy of training set (%)	Accuracy of testing set (%)
Subdivision grid method based on suboptimal solutions	$[2^1, 2^2]$	$[2^5, 2^6]$	0.01	C = 2, g = 61.3929	98.3333	96.6667
	$[2^3, 2^4]$	$[2^3, 2^5]$	0.01	C = 8, g = 20.9663	93.3333	90
	$[2^5, 2^6]$	$[2^8, 2^9]$	0.01	C = 32, g = 256	98.3333	96.6667
	$[2^8, 2^{10}]$	$[2^4, 2^5]$	0.01	C = 256, g = 16	91.6667	86.6667

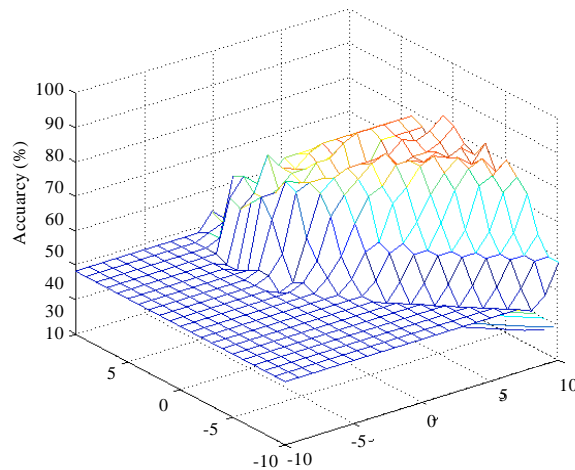


Fig. 1: Global search result based on grid search, Grid search method (3D view), Best C = 1024 g = 2

fact, the accuracy of base grid searching is not high indeed in Table 1. Whether subdivision grid can improve precision? After checking the efficient search region based on Fig. 1, reset the new search range: $c \in [2^0, 2^{10}]$, $g \in [2^{-5}, 2^{+10}]$ and reduce step length as 0.1. The second classification precision of testing set improves from 83.3333-86.6667% but value of c is 891.4441, still too large. Then the improved grid search method is applied in this paper. Shown as Fig. 1, there are some salient points in the searching area and their classification accuracies are very close to the highest accuracy by first search. The salient points are called as suboptimal solutions and they are found out by traversing all pairs of parameter. In this experiment, they can be expressed as data set: $(c, g) \in \{(2^1, 2^5); (2^2, 2^5); (2^3, 2^4); (2^5, 2^9); (2^6, 2^8); (2^9, 2^5)\}$. After

analyzing these six suboptimal solutions, four search ranges of the parameters are reset and the step length is set as 0.01. The regional search result is shown by Table 2. Finally the classification precision improves obviously and the value of c is small enough.

The experiment results illustrate that merely narrowing step length and enhancing the density of the grid search can not significantly improve the search accuracy. Shown by record 1, 2, 3 in Table 1, the step length is reduced from 1 to 0.01 and the search ranges are slightly adjusted, but the accuracy doesn't been highly improved while the calculation increases hugely. Similarly, if only subdivide the grid around the optimal solution; searching precision also doesn't necessarily improve obviously. Compared with record 1, 2, 4 in Table 1, the

Table 3: Classification accuracy comparison based on Subdivision grid (C = 891.444, g = 1.741)

Data set	SVM (%)	Weighted SVM (%)
Training set	86.67	90
Testing set	80	86.67

Table 4: Classification accuracy comparison based on suboptimal solutions (C = 2, g = 61.3929)

Data set	SVM (%)	Weighted SVM (%)
Training set	96.67	98.33
Testing set	93.33	96.67

search range gradually narrows around the optimal solution ($c = 1024, g = 2$); step length is changed from 1 to 0.1, then to 0.01, while the accuracy is still not changed obviously.

Many researches report that the testing accuracy will always be low if c value is too large because it leads SVM over learning. However, the suboptimal solutions' penalty factor value is relatively small, which can reasonably control the learning range of SVM and will not lead to over learning. This new method gain better prediction accuracy after subdividing grid around the suboptimal solutions for the regional search; shown as Table 2. In this example, the final parameter solution is (c, g) = (2, 61.3969).

Shown as Table 3 and 4, the classification accuracies are promoted after weighting the kernel function. Moreover, the improved grid searching method obviously increases the accuracy than traditional grid method.

CONCLUSION

This paper proposes the weighted SVM and improves the grid method by searching suboptimal solutions. The experiment confirms that the final solution is not the optimal solution supported by the first grid search but exists in the ranges of suboptimal solutions. Because the optimal solution has always been over trained, which leads to the value of penalty factors is too large. While make the second regional search around the suboptimal solutions can obtain the final solution whose accuracy is higher and penalty factors is smaller.

This paper also confirms that weighted SVM can distinguish the importance of different assessment indexes. It highlights strong characteristics' effect and reduces weak characteristics' influence in classification. It is very essential to reflect the effect difference of assessment indexes and reduce the classification mistakes. By the two aspects of improvements, the weighted SVM based on improved grid search is identified as an efficient classification method which is well adopted in flood disaster evaluation.

ACKNOWLEDGMENTS

This work was supported by the National Key Basic Research Program of China (973 Program, Grant No. 2007CB714107), the Special Research Foundation for the Public Welfare Industry of the Ministry of Water Resources (Grant No. 201001080, 201001034), and the Research Fund for the Doctoral Program of Higher Education of China (Grant No. 20100142110012). Multidimensional geological interoperable model experiment and modeling technology support (Gant No. 1212011120446).

REFERENCES

- Adankon, M.M. and M. Cheriet, 2010. Genetic algorithm-based training for semi-supervised SVM. *Neural Comput. Applic.*, 19: 1197-1206.
- Chang, C.C. and C.J. Lin, 2007. LIBSVM-A library for support vector machines. <http://www.csie.ntu.edu.tw/~cjlin/libsvm/index.html>.
- Chen, T., 2011. Selective SVM ensemble based on accelerating genetic algorithm. *Applic. Res. Comput.*, 28: 139-141.
- Cristianini, N. and J. Shawe-Taylor, 2000. *An Introduction to Support Vector Machines*. Cambridge University Press, Cambridge, UK.
- Jiang, H.Y. and J.M. Liu, 2012. Automatic liver cirrhosis diagnosis based on SSM and ACO-SVM. *J. Info. Comput. Sci.*, 9: 5927-5934.
- Jiang, H.Y., S.L. Jing, T.Y. Chai and X.J. Zhou, 2010. Sintering condition prediction based on SVM and PSO. *J. Northeastern Univ.*, 31: 494-497.
- Jiang, W.G., L. Deng, L.Y. Chen, J.J. Wu and J. Li, 2009. Risk assessment and validation of flood disaster based on fuzzy mathematics. *Progress Natural Sci.*, 19: 1419-1425.
- Kapp, M.N., R. Sabourin and P. Maupin, 2012. A dynamic model selection strategy for support vector machine classifiers. *Applied Soft Comput.*, 12: 2550-2565.
- Li, Q., J.Z. Zhou, D.H. Liu and X.W. Jiang, 2012. Research on flood risk analysis and evaluation method based on variable fuzzy sets and information diffusion. *Safety Sci.*, 50: 1275-1283.
- Lin, S.W., K.C. Ying, S.C. Chen and Z.J. Lee, 2008. Particle swarm optimization for parameter determination and feature selection of support vector machines. *Expert Syst. Applic.*, 35: 1817-1824.
- Ren, H. and M.D. Huo, 2009. Support vector machine optimized by particle swarm optimization algorithm for holding nail force forecasting. *Applic. Res. Comput.*, 3: 867-869.

- Shi, M.D. and C.C. Yang, 2008. Multiclass SVM-RFE for product form feature selection. *Expert Syst. Applic.*, 35: 531-541.
- Vapnik, V.N., 2000. *The Nature of Statistical Learning Theory*. 2nd Edn., Springer-Verlag, New York.
- Wang, J., H.Y. Du, X.J. Yao, Z. Hu, 2007. Using classification structure pharmacokinetic relationship (SCPR) method to predict drug bioavailability based on grid-search support vector machine. *Analytica Chimica Acta*, 601: 156-163.
- Wang, J.F., L. Zhang, G.X. Chen and X.W. He, 2012. A parameter optimization method for an SVM based on improved grid search algorithm. *Applied Sci. Technol.*, 6: 28-31.
- Wang, T.H., S.F. Tian and H.K. Huang, 2009. Feature weighted support vector machine. *J. Elect. Inform. Technol.*, 31: 514-518.
- Wang, Y.Y., S.C. Chen and H. Xue, 2011. Support vector machine incorporated with feature discrimination. *Expert Syst. Applic.*, 38: 12506-12513.
- Yuxia, H. and Z. Hongtao, 2012. Chaos optimization method of SVM parameters selection for chaotic time series forecasting. *Phys. Procedia*, 25: 588-594.
- Zhang, J.Y., S.L. Liu and Y. Wang, 2008. Gene association study with SVM MLP and cross-validation for the diagnosis of diseases. *Progress Natural Sci.*, 18: 741-750.