The Parameters Selection for SVM Based on Improved Chaos Optimization Algorithm*

Yong Wang, Yong Liu, Ning Ye, and Gang Yao

Institute of Intelligent Vision and Image Information,
China Three Gorges University, Yichang Hubei 443002, China
yongwang1971@sina.cn

Abstract. The parameters selection of support vector machine decides its study performance and generalization ability. The SVM model is greatly influenced by penalty factor C and the kernel function parameter such as σ for the radial basis function (RBF) kernel. To searching the best compound of parameters, a new algorithm is proposed based on improved chaos optimization strategy to realized automatic parameters selection for SVM. Chaos optimization algorithm is a global searching method in which the complexity and dimension of variables need not to be considered. Compared with the algorithms based on GA and PSO, the classification efficiency is improved greatly.

Keywords: Support vector machine, Chaos series, Chaos optimization, Parameter selection.

1 Introduction

Support vector machines (SVM) pioneered by Vapnik in 1992 is a new machine learning method which is systematic and properly motivated by statistical learning theory. Compared to the others learning algorithms such as neural network, SVM has characteristics of small sample machine learning and powerful generalization, which can effectively avoid overfitting, local minimum and curse of dimensionality [1]. The idea of SVM classifier is to use nonlinear transformation to translate nonlinear problems in original sample space to linear problems in high space. The nonlinear transformation is realized by defining a kernel function which satisfies Mercer theorem, so the performance of classifier is directly influenced by parameter selection of kernel. At the same time, the penalty factor C of SVM model which is a compromise between structural risk and sample error also influences performance of SVM. When using SVM, two crucial problems are confronted: how to choose the optimal input feature subset for SVM, and how to set the best kernel parameters σ and penalty factor C .At present, there is lack of mature theory to instruct the parameters selection which are acquired by experience, experiments and large space searching. The essence of parameters selection of SVM model is a process to solve combinatorial optimization problem. Its object is to choose optimal parameters to

* This work is partially supported by NSF Grant # 50805087 and # 60972162 to Yong Liu and Research Foundation for Talented Scholars of China Three Gorges University Grant #2010PY047 to Gang Yao.

J. Zhang (Ed.): ICAIC 2011, Part V, CCIS 228, pp. 376–383, 2011.

minimize the classification error rate of the SVM model [2]. In reference [3], an algorithm which is called Bilinear Grid Search Method (BGSM) to search optimal parameters compound in the parameters set. At first, the parameter C corresponding to high generalization ability for linear SVM is computed, then the parameter C is fixed, and the parameter σ of RBF kernel is searched corresponding to the highest generalization ability to obtain the compound (C , σ). Although the number of SVMs is decreased, the second linear searching stage relies heavily on the first stage parameter. The optimal parameters are computed by steepest descend method in reference [4], but the shortage is that this algorithm is sensitive to initial values. If the initialization is far from the optimal value, it is easy to fall into local optimization. An algorithm based on Genetic Algorithm(GA) to optimize the SVM parameters is presented in reference [5], and a more optimal parameters is acquired to use this method, but it is also easy to fall into local optimization and the iteration is large and it is time-consuming. A another algorithm based on particle swarm algorithm(PSO) is showed in literature[6] which is more efficient than GA and a global optimization is calculated, but the weakness is that early maturity appears.

Chaos optimization algorithm [7] is a global searching algorithm which is new and intelligent optimization strategy. The complexity and dimensions of variables is not considered when selecting parameters of SVM model. A method based on chaos optimization algorithm is showed to optimize the parameters of SVM model and the object function is the average minimum error rate which is computed by cross-validation. The improved chaos optimization algorithm is used to search the optimal combined parameters.

2 Model of SVM Parameter Optimization

A. SVM Model Analysis

SVM model transforms the learning process to an optimization issues according to the principle of minimize struck risk. To obtain smallest wrong divided sample and maximal margin ,that is optimal hyperplane, for training samples (x_i , y_i)(i= 1, 2,…,l, $x_i \in R^n$, $y_i \in \{1,-1\}^l$),SVM model needs to solve the problem that minimizes the object function(1) which meets the constraints of (2)

$$\phi(\omega,\xi)=\frac{1}{2}\|\omega\|^2+C\sum_{i=1}^{l}\xi_i \tag{1}$$

$$\text{s. t.} \quad y_i[(w^T \bullet z_i)+b]\geq 1-\xi_i\,;\ \xi_i\geq 0\ ,\ i=1,2,\ldots,l; \tag{2}$$

where training samples x_i is mapped to higher dimension space through function $Z_i=\phi(x_i)$, $\omega\in R^n$ is coefficient vector of hyperplane, $b\in R$ is threshold, ξ_i is slack variable, C ($C>0$) is penalty factor for samples which is divided wrongly.

Mentioned problem is transformed its dual problem through using Lagrange polynomial, which is searching Lagrange multipliers to maximize the object function as following:

$$Q(\alpha)=\sum_{i=1}^{l}\alpha^{i}-\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}a_{i}a_{j}y_{i}y_{k}K(x_{i},x_{j}) \tag{3}$$

subject to:

$$\sum_{i=1}^{l}\alpha_{i}y_{i}=0\ ;0\leq\alpha^{i}\leq C\ \text{,i=1,2,…,}l \tag{4}$$

So the corresponding classification decision function is

$$f(x)=\text{sign}\{\sum_{i=1}^{m}\alpha_{i}^{*}y_{i}K(x_{i},y_{i})+b^{*}\} \tag{5}$$

where non-negative α_i^* is support vector . m(m<l) is the number of support vectors, b^* is threshold corresponding to α_i^*, $K(x_i,y_i)=\phi(x_i)\bullet\phi(x_j)$ is kernel function which satisfies Mercer theorem. RBF kernel function is used in this paper which is $K(x_i,y_i)=\exp(-\|x_i-y_i\|)/\delta^2)$,where δ is the parameter of kernel.

B. Object Function of SVM Parameter Optimization

The generalization of SVM classifier is greatly affected by the parameters of C and δ .When C is very small, underfitting appears, but when C very large, overfitting appears that makes generalization worsened. When δ is very small, overfitting appears and δ is very large, underfitting appears. So the parameter δ is crucial to the generalization. In this paper, the error rate E_ζ of classifier is used to estimate the performance of SVM model. So the object is to minimize the E_ζ, which is selecting a optimized compound parameters to minimize the E_ζ . Let domain of penalty factor C and δ is respectively [a_1 , b_1] and [a_2 , b_2], so the optimization issues is

$$\begin{aligned}&\text{Min}\quad E_{\xi}(M)\\&\text{s. t. } M=\{C,\sigma\}\quad a_1\leq C\leq b_1\ ,a_2\leq\sigma\leq b_2\end{aligned} \tag{6}$$

C. The Analysis of Shortage of Presented Parameters Optimization of SVM Model

SVM model has good learning ability and generalization, but its performance depends on the selection of the parameters. Some scholars have studied and discussed from different angles for the parameter selection problem, which are divided three main methods. The first one is that users select the model parameters based on experience; the second one, which is the commonly used method currently, is cross-validation method; the third one is using modern intelligent optimization algorithm, such as the parameters selection based on GA, PSO and SA. The above three mentioned methods also have some disadvantages, which included too many human factors, requirement of the function which is continuously differentiable, long computing time and local minimum. Those shortcomings impact on the practical application of SVM seriously. Therefore, an improved chaos optimization algorithm which can be used to optimize the parameters of SVM model is presented in this paper.

3 The Parameter Optimization Based on Chaotic Series

A. The Analysis on Characteristics of Chaotic System

By taking advantage of ergodicity and random of chaotic motion, that is chaotic motion can traverse all states, Li Bing et al.[8]first proposed a chaos optimization algorithm (COA).The basic idea is that chaotic states is introduced to optimization variables like carrier wave with Logistic mapping ($x_{n+1} = \mu x_n(1-x_n)$).The traversing range of chaotic motion is extended to value range of optimization variables and then all fields is searched by chaos variables, that is used to solve optimization problem which is continuous and complicated. The essence of parameter optimization is a process of optimization search and this problem is always multi-peak that exists a plurality of local extremum. The ergodicity of chaos can avoid falling into local minimum in searching process, so chaos optimization algorithm is more efficient to obtain optimization parameters by means of its powerful global searching ability.

B. The Improved Chaos Optimization Algorithm Oriented SVM Parameter Optimization

Chaotic motion can traverse all state space. The larger the space is, the longer the traversal time. It is necessary to consider the reduction of search space of variable. Before narrowing the search space, traversal search needs to conduct many times in order to ensure that the optimal solution is in the narrowed search space. The search times are difficult to determine. Generally, it relates to the complexity of optimization function and the size of search space. So the search times just can be determined according to specific application. This reduces the versatility of algorithm. At the same time, because the termination condition is not given, the search is full of blindness and increases the amount of computation [9]. In order to reduce the blindness of search and improve the efficiency of search, the proposed improved and accelerated chaos optimization algorithm has the following advantages:

1) By introducing two-chaotic optimization mechanism, enhance the adequacy of the search, reduce the times of blind search and promote the efficiency of search. Chaos mapping is selecting as following:

$$x_{k+1} = \mu x_k(1-x_k) \tag{7}$$

and

$$y_{k+1} = \alpha y_k^3 - \alpha y_k + y_k \tag{8}$$

where (7) is logistic (μ =4); (8) is cube mapping which is chaos mapping when $\alpha \in$ [3.3, 4]. However, the possibility of the random value obtained throughout the entire search space is larger when α is a little bigger. So 3.9 is assigned to α .

2) The second improve for chaotic optimization search formula. The original chaos optimization algorithm searches only in the unilateral neighbourhood of Logistic map. When μ is assigned to 4, according to inequality 0< x^i <1, the product of adjustment coefficient α^i and chaotic variable x_i is greater than zero (α^i x_i >0) or less than zero (α^i x_i <0). It results in that the optimal solution x^* is searched in the unilateral neighbourhood. In this paper, a modified formula is proposed. It avoids the disadvantage of unilateral search. The modified formula is expressed as

$$x_i^r = x_i^* + \alpha_i(2x_i^r - 1)R \tag{9}$$

where x_i^r, r are chaotic variable and iteration times of the second chaos optimization search respectively, $r = 1, 2, \cdots$; x_i^* is the optimal variable for the current; Here the introduction of pseudo-random number R can be further reduced the search range and avoid the further deviation of optimal solution caused by the artificial introduction of adjustment factor, and speeds up the convergence.

C. The Steps of Improved Chaos Optimization Algorithm

Step 1: Initialize. According to (7) and (8), generate n initial chaotic variables with different trajectories.

Step 2: Map x_k^i, y_k^i to the range of the optimization variables respectively, have $mx_i^k = a_i^r + x_i^k(b_i^r - a_i^r)$, $my_i^k = a_i^r + (1/2)(y_i^k + 1)(b_i^r - a_i^r)$.

Step 3: Search using chaos variables. If $f(mx_i^k) < f_x^*$, have $f_x^* = f(mx_i^k)$, $mx_i^* = mx_i^k$; If $f(my_i^k) < f_y^*$, have $f_y^* = f(my_i^k)$, $my_i^* = my_i^k$; Else continue the search.

Step 4: Assign k+1 to k : $k = k + 1$, have $x_i^k = \mu x_i^k(1 - x_i^k)$; $y_i^k = \alpha(y_i^k)^3 - \alpha y_i^k + y_i^k$

Step 5: Repeat step 2 and 3. When the repeating times is more than h, judge whether the inequality $\|mx^* - my^*\| < \gamma\|a^r - b^r\|$ sets up, where $\gamma \in (0, 0.25)$ and h can be assigned to experienced 2000. If the inequality does not set up, repeat step 2 and 4 until this inequality meets. Then continue the following steps.

Step 6: Narrow the search range of each variable according to the following equations

$$a_i^{r+1} = \min(mx_i^*, my_i^*) - \xi\gamma\|mx_i^* - my_i^*\|;$$
$$b_i^{r+1} = \max(mx_i^*, my_i^*) + \xi\gamma\|mx_i^* - my_i^*\|.$$

where $\xi \in [1,2]$. If $a_i^{r+1} < a_i^r$, then $a_i^{r+1} = a_i^r$; if $b_i^{r+1} > b_i^r$, then $b_i^{r+1} = b_i^r$. The two conditional statements ensure the new range is not out of bound.

Step 7: $a_i^r = a_i^{r+1}, b_i^{r+1} = b_i^r, r = r + 1$;

Step 8: Return to step 2 until the best solution is found.

Step 9: Do the second carrier wave according to the equation $x_{i,n+1}^n = x^* + \alpha_i x_{i,n+1}^{'}$, where $x_{i,n+1}^{'}$ is chaos variables within the interval [-1,1]; α_i is adjustment coefficient; x^* is optimal solution for the current.

Step 10: Continue the iterative search using chaos variables after the second carrier wave. If $f(mx_i^k) < f_x^*$, then $f_x^* = f(mx_i^k)$, $mx_i^* = mx_i^k$.

Step 11: If the termination condition is met, terminate the search and output the optimal solution x^*, f^*; otherwise return to step 9.

The flow chart of above algorithm is showed in figure 1.

4 Simulation Experiment

In order to validate the effectiveness of the proposed algorithm, experiments are conducted on four classical classification data sets in UCI machine learning database. The parameter selection methods based on GA and PSO are used to compare with the proposed algorithm. The experimental results are displayed as table 1.

As shown in table 1, compared with the mentioned two methods, the computational time of the proposed method is greatly reduced under the same classification accuracy. It shows that the proposed improved chaos optimization algorithm for parameter selection not only improves the classification accuracy, but also reduces the computational time.

Table 1. Comparison of three methods

UCI data set		wine	Breast cancer	Heart disease	Australia credit card
The proposed algorithm	Time(s)	8.204	29.2291	37.5301	113.4846
	(c,g)	(195.155,4.37)	(33.512),2.162)	(144.145,0.033)	(174.8745,0752)
	Classification accuracy	99.8%	100%	99.7%	100%
The method based on PSO	Time(s)	10.8765	63.2585	47.0720	142.5834
	(c,g)	(126.47,3.9)	(46.177,3.871)	(136.068,0.291)	(177.392,0.504)
	Classification accuracy	98.8%	99.4%	97.9%	99.5%
The method based on GA	Time(s)	31.7380	188.4100	177.8151	608.1626
	(c,g)	(58.785,4.601)	(12.177,1.094)	(67.079,0.342)	(49.025,0.537)
	Classification accuracy	99.2%	98.3%	97.3%	98.5%

(c,g) is the optimization of compound (C , σ).

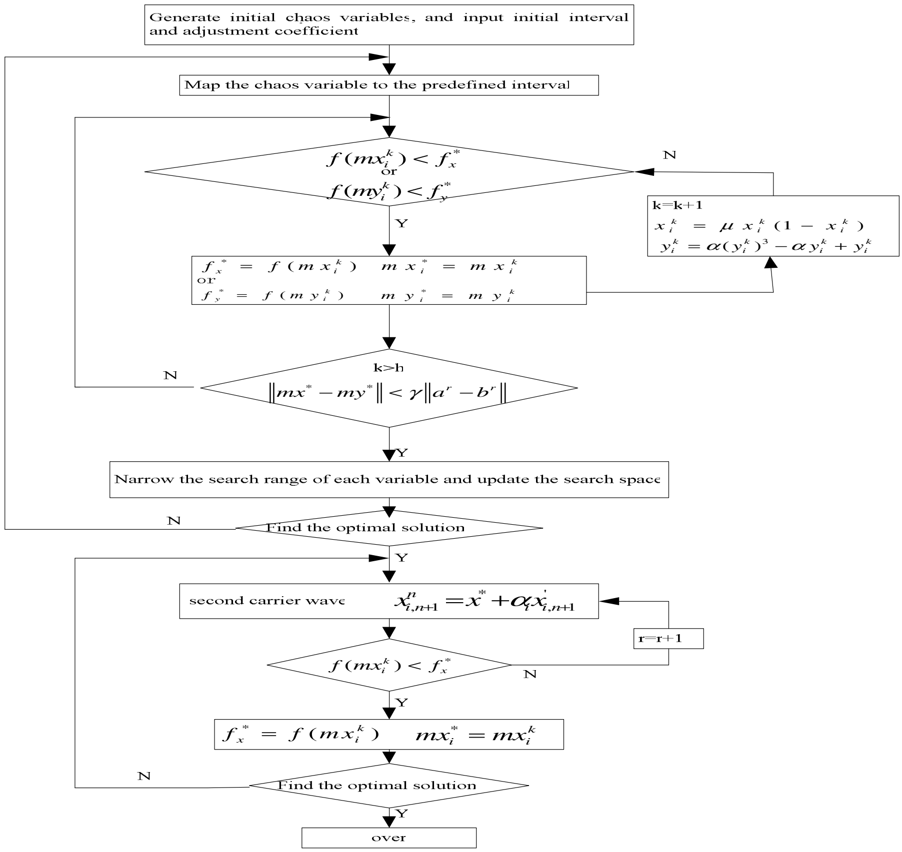

Fig. 1. The flow chart of chaos optimization algorithm

5 Conclusions

The performance of the SVM based on the structural risk minimization principle greatly depends on the parameters selection. Usually, the parameters are selected empirically and it limits the application of the SVM. In this paper, the improved chaos optimization algorithm is introduced into the parameter selection for SVM. The results of experiments demonstrate that the proposed improved chaos optimization algorithm for parameter selection method improves the generalization ability of the SVM, and make that the parameters selection does not depend on experience. It points out a new way for the application of the SVM.

Acknowledgment. This research has been supported by the Natural Sciences Foundation of China (NSFC) (Grant No. 50805087 and 60972162) and the Graduate Innovation Foundation of China Three Gorges University (Grant No.200914) and Research Foundation for Talented Scholars of China Three Gorges University (Grant No.2010PY047). It also has been done under the help of the team of the Institute of Intelligent Vision and Image Information.

References

1. Vapnik, V.N.: An overview of statistical learning theory. IEEE Trans. Neural Networks 10, 988–999 (1999)
2. Cherkassky, V., Ma, Y.Q.: Practical selection of SVM parameters and noise estimation for SVM regression. Neural Networks 17(1), 113–126 (2004)
3. Keerthi, S.S., Lin, C.J.: Asymptotic behaviour of support vector machines with Gaussian kernel. Neural Computation 15(7), 1667–1689 (2003)
4. Keerthi, S.S.: Efficient tuning of SVM hyperparameters using radius/margin bound and iterative algorithms. IEEE Transactions on Neural Network 13(5), 1225–1229 (2002)
5. Frohlich, H., Chapelle, O., Scholkopf, B.: Feature selection for support vector machines by means of genetic algorithms. In: Proceedings of the 15th IEEE International Conference on Tools with Artificial Intelligence (ICTA 2003), November 3-5, pp. 142–148. IEEE Computer Society, Los Alamitos (2003)
6. Ma, C., Ruan, Q.Q.: Parameter Selection for SVM Based on Discrete PSO. Computer Technology and Development 12, 20–23 (2007)
7. Zhang, H., He, Y.: Comparative study of chaotic neural networks with different models of chaotic noise. In: Wang, L., Chen, K., S. Ong, Y. (eds.) ICNC 2005. LNCS, vol. 3610, pp. 273–282. Springer, Heidelberg (2005)
8. Li, B., Jiang, W.: Chaos Optimization Method and Its Application. Control Theory & Applications 14(4), 613–615 (1997)
9. Xiu, C., Liu, X.: Optimization algorithm using two kinds of chaos and its application. Control Theory & Applications 18(6), 724–726 (2003)