Available online at http://www.mecs-press.net/ijem

A New Support Vector Machine Optimized by Simulated Annealing for Global Optimization

Jiayang Wang[a], Wensheng Wang[b], Shaogui Wu[c]

[a,b]College of Water Resources and Hydropower, Sichuan University, Chengdu 610065, China
[a]University of Information Technology, Chengdu 610041, China
[c]College of Chemistry and Material Science, Sichuan Normal University, Chengdu 610068, China

Abstract

SA-SVM model was proposed in which parameters were optimized by simulated annealing. Parameter σ (the kernel function) and C (the error discipline) are the key factors to the precision of SVM. Simulated annealing was used to optimize the key parameters of SVM to make enhancement on the forecasting effect of SVM. By applying this proposed model for several function optimizations, results of which demonstrate the improvement of SA-SVM on the high model accuracy in the optimization searching, and it can overcome the blindness of the model parameters.

Index Terms: support vector machines; simulated annealing; global optimization; parameters optimization

1. Introduction

SVM was developed by Boser et al[1] and Vapnik[2] to improve the accuracy of classifiers in machine learning and pattern recognition. It has the advantages of fast learning, global optimization and strong promotion compares with the traditional neural network methods which is based on empirical risk minimization. It can not only obtain the complex mapping relationship between the dependent variable and independent variables, but also get the evaluation and prediction results significantly better than other methods of pattern recognition and regression[3]. However, as a core element in the statistical learning theory, support vector machine's application is far from the desired theoretical effect. Parameter selection problem is a key factor to their application. In order to give full play to the performance of SVM, appropriate kernel function parameters must be chosen appropriately. Studies found that the kernel function parameter σ and the error penalty factor *C* are the key to the performance of SVM[4]. However, due to some limitations in practical applications, their generalization performance is sometimes far from the expected level. Selecting the appropriate kernel function

parameter σ and the error penalty factor C is essential for the performance of learning machine[4]. Therefore, SA-SVM was proposed based on the parameter optimization of SVM by simulated annealing.

2. The basic principle of SVM

Support Vector Machines is a powerful methodology for solving problems in nonlinear classification, function estimation and density estimation which has also led to many other recent developments in kernel based methods in general. The advantage of SVM is that it can achieve high accuracy with relatively small training sets. The details of SVM are discussed in article[2, 5, 6].

Set the training sample $\{ x_k , y_k \}^l$, where, $x_k \in R^n$ is n-dimensional input sample, $y_k \in R$ is the output sample. The approximation problem of the function is to find a corresponding function f for samples other than x after training through the sample. SVM estimates the unknown function by the following equation[4]:

$$f(x) = w \cdot \phi(x) + b \tag{1}$$

where, $\varphi(x)$ presents the mapped high dimensional feature space, b is the deviation value. The following formulae can be obtained by minimizing the risk function to coefficient w and b:

$$\min \left(\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l} |y - f(x)|_\varepsilon \right) \tag{2}$$

where, the first item $\frac{1}{2}\|w\|^2$ is called the model complexity item, the second item is the empirical error term determined by insensitive. SVM method transforms the above problems to the following antithesis question through introducing dot product nuclear function and using the Wolfe antithesis skills. Expressed as:

$$\max_{\alpha,\alpha^*} J = -\frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)k(x_i,x_j) - \varepsilon\sum_{i=1}^{l}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{l} y_i(\alpha_i - \alpha_i^*)$$

$$\text{s.t} \quad \begin{cases} \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0,C] \end{cases} \tag{3}$$

where, C ($C > 0$) is the penalty constant. Finally, the corresponding regression function (1) can be directly expressed as follows:

$$f(x) = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)k(x_i,x) + b \tag{4}$$

$k(x_i,x)$ is the nuclear function, which satisfy the Mercer condition. The study selects the Gauss nuclear function:

$$k(x,x_i) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2} \right) \tag{5}$$

3. The basic principle of SA-LSSVM

3.1. Parameter analysis of SVM

The basic principle of SVM is to map data in the input space to a high dimensional feature space by using a nonlinear mapping. Then, a linear mapping is made in the high dimensional space.

The width parameter of RBF precisely definite the structure of non-linear mapping functions in the high-dimensional feature space, which controls the complexity of the final solution to affect the linear regression of the high dimensional feature space. When decreases, the complexity of the regression function increase, and easy to cause over-fitting, the value of is too large or too small to cause the deterioration of the generalization performance of the system. So proper selection of is the key to the generalization ability of the regression model. The penalty factor C can adjust between the training error and the model complexity, indirectly affect the generalization ability of the regression model. The generalization of the system performance will be deteriorated when C is too large or too small.

Several methods have been developed for choosing the best hyperparameter values[7]. The empirical formula is calculated by Wenjian Wang et al[8] according to the fisher discriminant function, which gives the calculated values in the regression method according to the scale space theory. It can provide a channel for the optimal choices. Staelin[9] used the design of experiment method (DOE) to get the appropriate parameter values, which can significantly reduced the search costs.

3.2. simulation annealing

The simulation annealing[10] is one kind of global intelligence optimization algorithms, whose principle is the simulation of metal annealing process. In the optimum searching process of t he simulation annealing algorithm, it accepts both the optimal solution and the sub-optimal solution with a certain probability. This enables it to search optimal solution in more new partial regions of t he solution space. The probability is related to the initial temperature of the annealing process, and the higher the temperature, the greater the probability of receiving a sub-optimal solution. The possibility will finally vanish along with the gradual reduction of the temperature. Due to this mechanism, simulation annealing algorithm has better capability of avoiding the partial extreme value compared to other optimization algorithms

3.3. The basic principle of SA-SVM

SA is used to optimize the parameters of SVM (C , σ), The generic flow of SA-SVM is given as follows:

Step 1: initialization: Given initial point x_0, the initial temperature, final temperature, annealing form, within the same temperature cycles, etc.

Step 2: Calculation the fitness value. Divide a small portion (generally the 10% of all training samples) of the sample data used for training as the validation sample set. Use the remaining training samples to training the SVM model to fit the requirements. Predict the validation samples after obtaining the parameter values of the SVM model.

Step 3: random disturbance, get a new point $x' = x + \Delta x$, updates the size of the advantages and fitness with optimal operation of SA.

Step 4: cooling.

Step 5: Check the termination condition, if satisfied, end the optimization; otherwise $t = t + 1$, go to step 2. End of the search for optimal conditions for termination of the temperature or the maximum evolution generation;

Step 6: the optimal value of the SA algorithm is the optimal parameter vector (C, σ) in the SVM. Learn all the training samples get all the training samples of simulation accuracy by the optimized SVM.

4. Function Fitting Analysis Based on SA-SVM

The reliability and efficiency of SA-SVM are tested for several multi-optima functions to test the calculation accuracy, learning ability and the model accuracy of it. Multi-peak positive function, Schaffer function and the function are shown respectively below.

Example 1: minimum of Camel function.

$$f(x_1, x_2) = \left(4 - 2.1x^2 + \frac{x_1^4}{3}\right)x_1^2 + x_1 x_2 + \left(-4 + 4x_2^2\right)x_2^2 \tag{6}$$

where search domain is $-2 \leq x_1, x_2 \leq 2$.

Example 2: maximum of Multi-peaks function.

$$\max(y_1) = e^{-0.001x} \cos^2(0.8x) \qquad x \in [0, \ 18] \tag{7}$$

The theoretical maximum of positive function is: $y = 1.0000$ at $x = 0.0000$. The function has 4 peak values except the maximum: $y(x) = 0.9961$, $x = 3.9262$, $y(x) = 0.9922$, $x = 7.853 \, 2$, $y(x) = 0.9883$, $x = 11.7810$, $y(x) = 0.9844$, $x = 15.7090$.

Example 3: minimum of the function.

$$f(x_1, x_2) = x_1^2 + 2x_2^2 - 0.3\cos(3\pi x_1) - 0.4\cos(3\pi x_2) + 0.7 \tag{8}$$

where search domain is $-10 \leq x_1, x_2 \leq 10$.

For all the mentioned function, calculations have been performed to study the reliability of SA-SVM. 1000 points were generated within the range of variable values as the training samples for each function. The penalty constant C and the kernel function σ of SVM were optimized by SA. Then 100 different points were generated within a range of function interval as the fitting sample. Results and error analysis were shown in Table 1.

Table 1. Results and error analysis

function	C	σ	Training Error	Fitting Error
Camel function	181.7749	0.0073	9.5327e-004	0.0053
Multi-peaks function	94.7959	0.0178	3.8823e-008	1.2731e-007
function	152.7678	0.0097	0.0201	0.1029

The fitting results of the three typical functions were shown in Fig. 1. It can be seen from Fig. 1 that the function fitting values by SA-SVM mode were very close to the theoretical value, SA-SVM mode has a strong learning ability and a high precision.
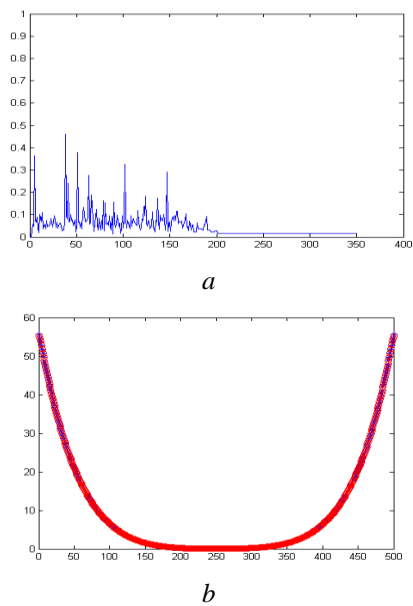
a

b

Fig. 1. Training error (*a*) and the fitting figure (*b*) of the Camel function
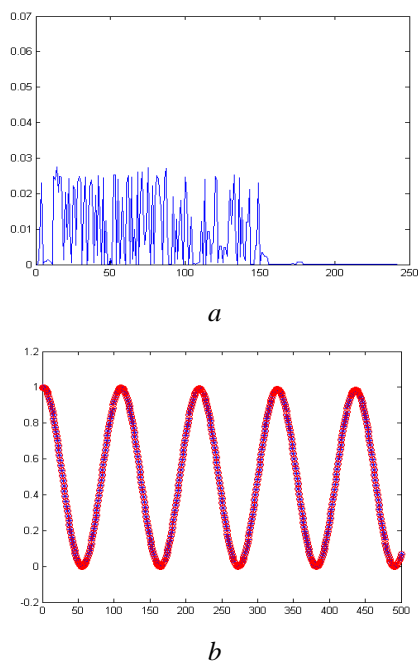

a

b

Fig. 2. Training error (*a*) and the fitting figure (*b*) of the multi-peaks function.
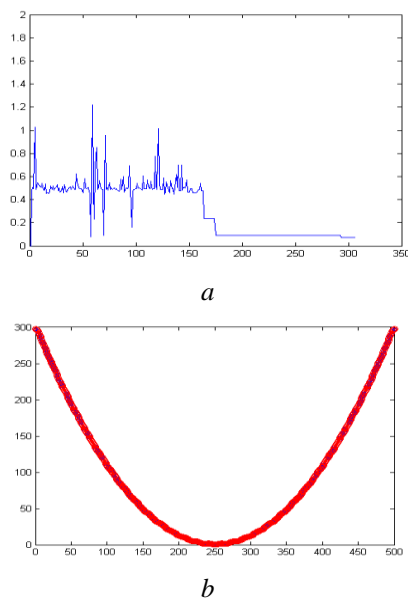
Fig. 3. Training error (*a*) and the fitting figure (*b*) of the third function.

5. Conclusions

(1) SA algorithm is used to optimize the parameters of SVM, to overcome the blindness of the model parameters, and present widely used spreadsheet through cross-validation method to determine the parameters, compared with the TS algorithm optimization parameters have a more clear theoretical guidance.

(2) The article only adopted RBF as the kernel function, but in fact different kernel function selections have direct effect to SVM model. Therefore in-depth studies on how to choose other type of kernel function will be needed in future.

Acknowledgment

References

[1] B. E. Boser, I. M. Guyon, and V. Vapnik, "A Training Algorithm for Optimal Margin Classifiers," presented at Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory, 1992."
[2] V. V. N, *The Nature of Statistical Learning Theory*. NY: Springer-Verlag, 1995.
[3] X. Y. Wang, H. Y. Yang, and C. Y. Cui, "An SVM-based robust digital image watermarking against desynchronization attacks," *Signal Processing*, vol. 88, pp. 2193-2205, 2008.
[4] L. I. Xiang, Y. Shangdong, and Q. Jianxun, "A new support vector machine optimized by improved particle swarm optimization and its appl ication," *J. CENT. SOUTH UNIV. TECHNOL.*, vol. 13, pp. 568-572, 2006.

[5] C.-h. Li, Z.-d. Lu, and K. Zhou, "SVR-parameters selection for image watermarking," presented at Proceedings of the 17th IEEE International Conference on Tools with Artificial Intelligence (ICTAI' 05), Hong Kong, China, 2005.

[6] C. N, S.-T. J, et. al, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. New York, 2000.

[7] B. O and B. V. L, "On Domain Knowledge and Feature Selection Using a Support Vector Machine," *Pattern Recognition Letters*, pp. 475-484, 1999.

[8] W. Wang, Z. Xu, W. Lu, and et. al, "Determination of the spread parameter in Gaussian kernel for classification and regression," *Neurocomputer*, vol. 55, pp. 643-663, 2003.

[9] S. Carl, "Parameter selection for support vector machines." 2005. http://www.hpl.hp.com/techreports/2002/HPL-2002-354Rl.html.

[10] N. Azizi and S. Zolfaghari, "Adaptive temperature control for simulated annealing: a comparative study," *Computers & Operations Research*, vol. 31, pp. 2439-2451, 2004.