

# Chaos Particle Swarm Optimization Algorithm for Optimizing the Parameters of Support Vector Machine

Zi-de Tian

Mathematics Department, Baicheng Normal College, Baicheng, 137000, China

**Keywords:** Support vector machine; Chaos particle swarm optimization; Parameters optimization.

**Abstract.** SVM parameter selection determines the performance of its learning and generalization ability. As many in the choice of the number of selectable parameters, blindly searching optimal parameters in a number of parameters are required tremendous time cost, and it is difficult to obtain the optimal parameters. To solve the support vector machine parameter optimization problem, a chaos particle swarm optimization algorithm is introduced to determine the parameters of SVM. Chaos theory is applied in PSO algorithm to improve the diversity of the population and particle traversal search, which can effectively improve the PSO algorithm convergence speed and accuracy, and SVM mode is optimized. Through a specific example, its results demonstrate that the chaos PSO has a good high efficiency and higher accuracy than the traditional PSO applied in SVM classifier.

## Introduction

Support vector machine (SVM) is based on the principle of minimizing the structural risk methods. The algorithm is a convex quadratic optimization problem, which find the global optimal solution. On the one hand, it can solve the practical problems of small sample, nonlinearity, curse of dimensionality, etc. [1-3] On the other hand, it can also overcome the defects including converges slow and easy to fall into local minimum of neural network. However, SVM has a prominent problem in a specific application process, which is how to select the key parameters that affect the performance of the algorithm, since the parameter selection of SVM determines the performance of its learning and generalization ability. Currently, the general method to select the parameters concludes gradient descent algorithm, genetic algorithm and particle swarm optimization (PSO). But these methods have their drawbacks [4-7]. In view of the chaos optimization characters concluding randomness, ergodicity, sensitivity to initial conditions, this paper introduce chaos optimization theory into particle swarm optimization algorithm. A chaos particle swarm optimization (CPSO) algorithm for optimize the parameters of SVM is proposed. Chaos performance is used to improve the ergodicity of diversity and particle search of population, which can improve the efficiency of the algorithm and the classification accuracy of SVM classifier[8-11].

## Support vector machine

SVM is a new learning method based on structural risk minimization principle of the machine, which can make full use of the limited sample learning acquisition decision function with high generalization ability. Considering a two classification models, a training sample is defined as  $\{(x_i, y_i)\}$ , where  $x_i$  represents the input vector and  $y_i$  represents the classification sign. The two classes of linear discriminant function of separable cases are as follows:

$$f(x) = w \cdot \varphi(x) + b \quad (1)$$

Where  $x$  is the sample vector,  $w$  is the weight vector,  $b$  represents the classification threshold. This process can be described as Fig 1.

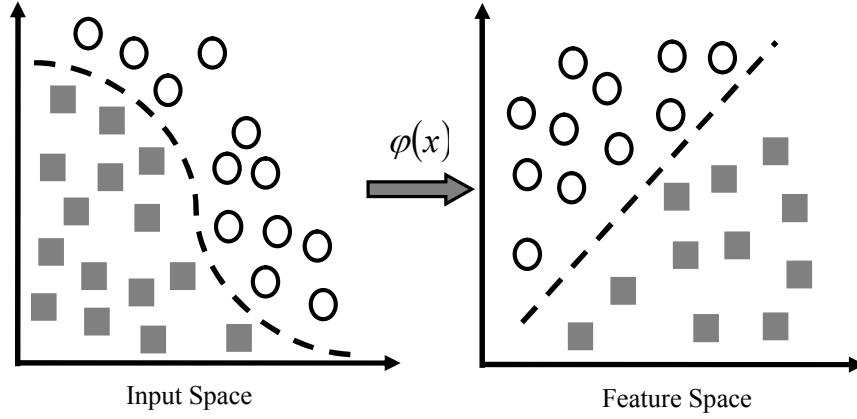


Fig 1: Illustration of the transformation process of SVM model

Supposing there is a classification decision plane:

$$f(x) = w \cdot \varphi(x) + b = 0 \quad (2)$$

Which satisfies:

$$f(x) = \begin{cases} w \cdot \varphi(x) + b > 0 & y_i = +1 \\ w \cdot \varphi(x) + b < 0 & y_i = -1 \end{cases} \quad (3)$$

Thus, Eq.(2) is defined as the classification and ultra flat surface of support vector.

A two types of linear separable linear sample set  $\{(x_i, y_i)\}^N, i = 1, 2, \dots, n$  is given. The weight vector  $w$  in hyperplane and classification threshold  $b$  are both obtained. As to any support vector  $x_{sv}$  and its classification label  $y_{sv}$ , it satisfies the following condition.

$$f(x_{sv}) = W^T x_{sv} + b = \begin{cases} +1 & y_{sv} = +1 \\ -1 & y_{sv} = -1 \end{cases} \quad (4)$$

The maximum interval of the classification plane can be calculated as  $\frac{2}{\|w\|}$ , Therefore, the maximum class interval is equivalent to the minimum. If the surplus plane need to classify all samples correctly, it must satisfy:

$$y_i (W^T x_i + b) - 1 \geq 0, i = 0, 1, 2, \dots, N \quad (5)$$

Thus, the SVM hyperplane by solving the following constrained optimization problem was the solution:

$$\min_w \frac{1}{2} W^T W \quad (6)$$

Which is a A typical quadratic programming problem. The optimal solution is a saddle point of the following Lagrangian:

$$L(w, b, a) = \frac{1}{2} (w, w)^T - \sum_{i=1}^l \alpha_i [(x_i w - b) - 1] \quad (7)$$

Where  $a_i$  is the Lagrange multipliers, The above optimization problem inverts into a dual form to get the solution:

$$\max W(\alpha) = \min \left( -\frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j k(x_i, x_j) \right) \sum_{i=1}^n \alpha_i \quad (8)$$

According KKT conditions, those points on the boundary of two types of sample points fall in support vector machine SVM required solution. Classification decision function is obtained as follows:

$$f(x) = \text{sign} \left( \sum_{i,j=1}^n \alpha_i y_i k(x_i, x_j) + b \right) \quad (9)$$

Where  $k(x_i, x_j)$  is the kernel function.

The Gaussian kernel function is selected as the support vector machine kernel function in this paper. The form of Gaussian kernel function is shown as follows:

$$K(x, y) = \exp \left( -\frac{\|x - y\|^2}{\sigma^2} \right) \quad (10)$$

Thus, the optimal separating hyperplane is required:

$$f(x) = \text{sign} \left( \sum_{i,j=1}^n \alpha_i y_i \exp \left( -\frac{\|x_i - x_j\|^2}{\sigma^2} \right) + b \right) \quad (11)$$

For SVM with Gaussian kernel function, the parameters conclude adjustment parameter  $C$  and core width  $\sigma$ . The parameter  $C$  is made between the structure and the risk of sample error compromise. The value of  $C$  is related to the tolerable error. A larger value allows small error and the smaller value allows larger errors. Core width  $\sigma$  is related to the input space of learning sample or the width. If the extent of the input sample is large, the value is large. On the contrary, if the extent of the input sample is small, the value is also small.

### Chaos particle swarm optimization algorithm

PSO algorithm is simple, but it has a weak local search capability, easy to fall into local optima, the late evolutionary slow convergence and other defects. Because of the chaotic motion with ergodicity, randomness, sensitivity to initial conditions and other characteristics, so chaos principle is introduced in PSO in order to enhance the diversity of the population, particle traversal search, improve the ability to get rid of local extreme points and convergence speed and accuracy.

This paper uses Logistic equation to construct chaotic sequence:

$$x_{n+1} = \mu x_n (1 - x_n), \quad (12)$$

Where  $\mu$  is the control parameter. When  $\mu$  equals 4, the system is fully in a chaotic state.

Using the position and velocity of particles chaotic sequence initialization neither change the nature of randomness to initialize the particle swarm optimization algorithm has, at the same time being able to put to good use to improve the diversity and chaos characteristics of particle populations search ergodic, generating on the basis of a large number of initial population, merit elect initial groups.

In the current search to the entire particle swarm optimal location based on chaotic sequence to generate a new, substitute the current position of a particle in the particle swarm chaotic sequence with the optimal position of the particle. Introducing chaotic sequence search algorithm generates a number of local optimum iteration points in the neighborhood, in order to help the inert particles to escape from the local minimum point, to quickly find the optimal solution.

Research indicates that although the chaotic search is of ergodicity and may avoid from local optima, but a simple chaos search needs a large number of loop steps to obtain a better solution and the result is sensitivity to the initial value. Combining chaos optimization algorithm and PSO algorithm, we proposed chaotic particle swarm optimization (CPSO), in which PSO algorithm for global search, and Chaos Algorithm based on the results of the PSO algorithm for local search. Main steps of CPSO algorithm are as follows:

Step 1 Randomly initialize the particle state positions and velocities. They are usually generated randomly in the permitted range, each current state position of particles set its state position of local optima, and compute the corresponding local optima (individual objective function's fitness), while the global optima (global fitness value) is the best of all local optimas, recording the serial

number of the particle of the global optima, the current position of the global optima particle set its position of global optima;

Step 2 Compute the fitness value (the value of objective function) corresponding to each particle;

Step 3 For each particle, compare each fitness value with the local optima, if better, then update the current local optima state;

Step 4 For each particle, compare the fitness values with the global optima, if better, then update the current global optima;

Step 5 According to equation (1) and equation (2), update fly velocity and state position of each particle;

Step 6 According to equation (4) , generate the chaotic variables between [-B, B] (where B is upper limit of disturb range) to disturb the position of updated particles ;

Step 7 Search in state space and compare the new state (the value of objective function) with the previous one , if better, then update the current state position;

Step 8 If the pre-set stop condition is met (usually the fitness value is less than a given value or the number of iteration loops is equal to or larger than the given maximum number N of iteration loops) and output final results , computation stops ; if not, return to step 2 to continue to carry out next loop of state search computation.

### Parameters optimization of SVM based on CPSO

The fitness function of CPSO is defined as the mean square error (MSE) which can directly reflect the performance of SVM. MSE is defined as follows:

$$MSE = \sqrt{\frac{\sum_{i=1}^n (y' - y_i)^2}{n}} . \quad (13)$$

Where  $y_i$  is the actual value and  $y'$  is the forecasting value.

The steps of CPSO-SVM is specifically described as follows:

(1) Parameter settings of CPSO. The parameters include the size of the particle populations, fitness error limit, the maximum allowed number of iterations, inertia weight and learning factors.

(2) Set the position and velocity of chaos particle. The initial velocity of each particle  $v_i$  is randomly generated.

(3) The personal best position of each particle is set as the current position. Calculate the fitness of each particle and take the fitness of the best individual particles corresponding to the extreme as the initial global extremes.

(4) Implement iterative calculation according to Eq.(14), Eq.(15) and Eq.(16). Then update the position and velocity of the particles.

$$v_{id} = wv_{id} + c_1r(p_{id} - x_{id}) + c_2R(g_d - x_{id}) . \quad (14)$$

$$x_{id} = x_{id} + \alpha v_{id} . \quad (15)$$

$$w = w_{\min} + \frac{(iter_{\max} - iter)(w_{\max} - w_{\min})}{iter_{\max}} . \quad (16)$$

Where  $v_i$  represents the particle velocity,  $w$  represents the inertia weight,  $iter$  represents the current iterations.

(5) All the best position are implement the chaotic optimization. Through iteration of Logistic equation, the chaos variable sequence  $x_i^m$  is generated. Then, the  $x_i^m$  returns to the original solution space through the inverse mapping. Thus, each feasible solution in the original solution space of chaotic variables can calculate the fitness value and get the best feasible solution performance  $p^*$  .

(6) Optionally substituted a particle location in the current population with  $p^*$  .

(7) Stop searching when it reaches the maximum number of iterations, otherwise it returns to step three.

### Specific example and results analysis

In order to measure the validity of chaotic particle swarm optimization SVM model, the 690 cases of a credit card company are used in this paper, in which each piece of data has 15 properties which contains 6 continuous attributes , 8 items type attribute and a target attribute. The credit card data is divided into 10 groups of simulation experiments. The comparative experiments is the traditional particle swarm optimization algorithm.

Chaos particle swarm optimization algorithm parameter settings are as follows: the size particle swarm population is 30, learning factor  $c_1$  and  $c_2$  are both 2, the maximum iteration times is 1000. Meanwhile, the two parameters  $C$  and  $\sigma$  in this paper implement B= binary coding. The searching scope of  $C$  is set from 0 to 1000, and the searching scope of  $\sigma$  is set from 0.1 to 100.

The stop condition of these two algorithms is as follows: When the maximum evolution algebra is over 1000 generations or the difference of the absolute value for 10 consecutive generations is less than 0.01, the optimization process is stopped. Then the parameters are the optimal parameters of SVM. Its classification results are shown in Table 1.

Table 1: Classification results

Groups	Correct rate of classification [%]	
	Traditional PSO	CPSO
1	85.5	86.4
2	76.8	77.6
3	76.8	77.6
4	86.9	87.8
5	91.3	92.2
6	86.9	87.8
7	85.5	86.4
8	88.4	89.3
9	82.6	83.4
10	86.9	87.8

Classification results in Table 1 indicate that, compared with traditional PSO method, when given the data classification, the chaos particle swarm optimization algorithm has a higher precision and its number of iterations reduces significantly.

### Conclusions

SVM classification accuracy largely depends on its chosen parameters, this paper applies chaos particle swarm optimization algorithm to solve the problems conclude artificially selected blindness and randomness in support vector machine, which ensure the accuracy of the classification results. Using the advantages of chaos optimization including randomness, ergodicity, sensitivity to initial conditions, etc. can overcome the shortcomings of local optimum value of PSO, improve the diversity of the population and particle traversal search, improve the convergence speed and accuracy.

### References

[1] Maksim Lapin, Matthias Hein, Bernt Schiele. Learning using privileged information: SVM+ and weighted SVM. *Neural Networks*, 53(11), pp. 95-108,2014.

- [2] Zhigang Wang, Zengshun Zhao, Shifeng Weng, Changshui Zhang. Solving one-class problem with outlier examples by SVM. *Neurocomputing*, 149 (4), pp. 100-105,2015.
- [3] Michal Pluhacek, Roman Senkerik, Donald Davendra, Zuzana Kominkova Oplatkova, Ivan Zelinka. On the behavior and performance of chaos driven PSO algorithm with inertia weight. *Computers & Mathematics with Applications*,66 (6),pp. 122-134,2013.
- [4] Mehdi Sadeghpour, Hassan Salarieh, Gholamreza Vossoughi, Aria Alasty. Multi-variable control of chaos using PSO-based minimum entropy control. *Communications in Nonlinear Science and Numerical Simulation*, 16 (18), pp. 2397-2404,2011.
- [5] WEN Fenghua, XIAO Jihong, HE Zhifang, GONG Xu. Stock Price Prediction based on SSA and SVM. *Procedia Computer Science*, 31 (7), pp. 625-631,2014.
- [6] Limeng Cui, Yong Shi. A Method based on One-class SVM for News Recommendation. *Procedia Computer Science*,31. (11), pp. 281-290,2014.
- [7] Benlan He, Yong Shi, Qian Wan, Xi Zhao. Prediction of Customer Attrition of Commercial Banks based on SVM Model. *Procedia Computer Science*, 31 (18), pp. 423-430,2014.
- [8] Jan Chorowski, Jian Wang, Jacek M. Zurada. Review and performance comparison of SVM- and ELM-based classifiers. *Neurocomputing*,128 (8), pp. 507-516,2014.
- [9] Liu Jian, Tan Tianyuan. LS-SVM based substation circuit breakers maintenance scheduling optimization. *International Journal of Electrical Power & Energy Systems*. 64, pp. 1251-1258, 2015.
- [10] Limeng Cui, Yong Shi. A Method based on One-class SVM for News Recommendation. *Procedia Computer Science*. 31, pp. 281-290, 2014.
- [11] Benlan He, Yong Shi, Qian Wan, Xi Zhao. Prediction of Customer Attrition of Commercial Banks based on SVM Model. *Procedia Computer Science*. 31, pp. 423-430, 2014.