Parameters Optimization of Least Squares Support Vector Machines and Its Application

Chunli Xie¹,²
1. Dalian University of Technology/ School of Electronic and Information Engineering, Dalian, 116024, China
2. Dalian Nationalities University/College of Electromechanical and Information Engineering, Dalian, 116024, China
Email: chunlix@dlnu.edu.cn

Cheng Shao¹ and Dandan Zhao³
3. Dalian Nationalities University/School of Computer Science and Engineering, Dalian, 116024, China
Email: cshao@dlut.edu.cn, zhaodd@dlnu.edu.cn

Abstract—Parameters optimization plays an important role for the performance of least squares support vector machines (LS-SVM). In this paper, a novel parameters optimization method for LS-SVM is presented based on chaotic ant swarm (CAS) algorithm. Using this method, the optimization model is established, within which the fitness function is the mean square error (MSE) index, and the constraints are the ranges of the designing parameters. After having been validated its effectiveness by an artificial data experiment, the proposed method is then used in the identification for inverse model of the nonlinear under-actuated systems. Finally real data simulation results are given to show the efficiency.

Index Terms—Least Squares Support Vector Machines, Parameters Optimization, Chaotic Ant Swarm Algorithm

I. INTRODUCTION

A novel statistical learning method called Support Vector Machines (SVM) was presented by Vapnik in 1995. Due to the advantages such as the complete statistical learning theory foundation and perfect study ability, SVM has become quite an active research field in machine learning and broadly used in many fields such as pattern recognition and regression estimation problems [1, 2]. The classical training algorithm of SVM is equivalent to solving a quadratic programming with linear and inequality constraints. Least squares support vector machines (LS-SVM) has been recently introduced by Suykens et al. as reformulations to standard SVM [3, 4], which simplifies the training process of standard SVM in a great extent by replacing the inequality constraints with equality ones. The simplicity of LS-SVM promotes the applications of SVM, and many pattern recognition and function approximation problems have been tackled with LS-SVM in the last decade [5-9].

The parameters in regularization item and kernel function are called parameters in LS-SVM, which play an important role for the algorithm performance. The existing techniques for tuning the parameters in LS-SVM can be summarized into two kinds: one is based on analytical techniques, the other is based on heuristic searches. The first kind of techniques determines the parameters with gradients of some generalized error measures [10]. And the second kind of techniques determines the parameters with modern heuristic algorithms including genetic algorithms (GA), simulated annealing algorithms (SA), particle swarm optimization algorithms (PSO) and other evolutionary algorithms [11-15]. Iterative gradient-based algorithms rely on smoothed approximation of a function. So, it does not ensure that the search direction points exactly to an optimum of the generalization performance measure which is often discontinuous. Grid search [16] is one of the conventional approaches to deal with discontinuous problems. However, it needs an exhaustive search over the space of parameters, which must be time consuming. This procedure also needs to locate the interval of feasible solution and a suitable sampling step.

In this paper, a novel algorithm of parameters optimization is presented based on the principles of the chaotic ant swarm (CAS) algorithm. Inspired by the chaotic and self-organizing behavior of the ants in nature, the novel CAS [17] algorithm is developed in 2006, which combines the chaotic behavior of individual ant with the intelligent foraging actions of ant colony via the organization variable for solving optimization problems. Similar to GA, EA and PSO, the CAS algorithm is a population-based optimization tool, which searches for optima by updating generations. However, unlike GA and EA, the CAS algorithm does not need evolutionary operators such as crossover and mutation. Compared to GA and EA, the advantages of CAS algorithm are that it possesses the capability to escape from local optima, is easy to be implemented, and has fewer parameters to be tuned. Compared to PSO, the advantages of CAS algorithm are that it has higher convergent precision. The CAS algorithm has been successfully applied to parameters estimation, artificial network training and fuzzy system control, etc [18-26]. The CAS algorithm is used to the parameters optimization of LS-SVM, and the
feasibility of this approach is examined on the testing function and nonlinear under-actuated systems.

This paper is organized as follows. The LS-SVM regression algorithm is briefly reviewed in Section 2. Parameters optimization algorithm based on the CAS algorithm is addressed in Section 3. The results of testing and simulation are presented to demonstrate the effectiveness of the proposed method in Section 4. The application of LS-SVM based on the CAS Algorithm is given in Section 5. Finally, the paper is concluded in Section 6.

II. LS-SVM REGRESSION

The LS-SVM, evolved from the SVM, changes the inequality constraint of a SVM into an equality constraint and forces the sum of squared error (SSE) loss function to become an experience loss function of the training set. Then the problem has become one of solved linear programming problems. This can be specifically described as follows [4]:

Given the following training sample set \( D \):

\[
D = \{(x_k, y_k) | k = 1, 2, \cdots, N \}
\]

where \( N \) is the total number of training data pairs, \( x_k \in R^n \) is the regression vector and \( y_k \in R \) is the output. According to SVM theory, the input space \( R^n \) is mapped into a feature space, and then the linear equation in the feature space can be defined as:

\[
f(x) = w^T \phi(x) + b
\]

where the nonlinear mapping \( \phi : R^n \rightarrow R^m \) maps the input data into a so-called high dimensional feature space (which can be infinite dimension). The regularized cost function of the LS-SVM is given as:

\[
\min J(w, e) = \frac{1}{2} w^T w + \frac{1}{2} y \sum_{k=1}^{N} e_k^2
\]

\[
s.t. y_k = w^T \phi(x_k) + b + e_k, \quad k = 1, 2, \cdots, N
\]

where, \( w \in R^n \) is the weight vector, \( e_k \in R \) is slack variable, \( b \in R \) is a bias term and \( y \in R \) is regularization item. The Lagrangian corresponding to Eq. (2) can be defined as follows:

\[
L(w, b, e, \alpha) = J(w, e) - \sum_{k=1}^{N} \alpha_k \left( w^T \phi(x_k) + b + e_k - y_k \right)
\]

where \( \alpha_k \in R \) \((k = 1, 2, \cdots, N)\) are the Lagrange multipliers. The KKT conditions can be expressed by

\[
w = \sum_{k=1}^{N} \alpha_k \phi(x_k)
\]

After elimination of \( w \) and \( e_k \), the solution of the optimization problem can be obtained by solving the following set of linear equations

\[
\begin{bmatrix}
    \alpha \\
    b
\end{bmatrix} =
\begin{bmatrix}
    0 & -I & 0 \\
    -I & \Omega + \gamma^{-1} I & -y \\
\end{bmatrix}^{-1} y
\]

with \( y = [y_1, \cdots, y_N] \) \( \in R^N \), \( \gamma = [1, \cdots, 1] \) \( \in R^N \), \( \alpha = [\alpha_1, \cdots, \alpha_N] \) \( \in R^N \) and \( \Omega \) is an \( N \times N \) kernel matrix. By using the kernel trick [2], one obtains

\[
\Omega_{kl} = \phi(x_k)^T \phi(x_l) = K(x_k, x_l), \forall k, l = 1, 2, \cdots, N.
\]

And the resulting LS-SVM regression model becomes

\[
f(x) = \sum_{k=1}^{N} \alpha_k K(x, x_k) + b
\]

III. PARAMETERS OPTIMIZATION OF LS-SVM BASED ON CAS ALGORITHM
A. Overview of CAS Algorithm.

Ants have attracted many scientists’ significant interests because their colonies can achieve the self-organizing behavior and the high level of structure. Most of the existing ant-inspired optimization algorithms are based on the random meta-heuristic of nondeterministic probability theory. However, Cole suggested that ant colony exhibits a periodic behavior while single ant show low-dimensional deterministic chaotic activity patterns [27]. From the view of dynamics, the chaotic behavior of single ant has some relation to the self-organizing and foraging behaviors of ant colony. The chaotic behavior of individual ant and the intelligent organization actions of ant colony are adaptations to the environment. These behaviors help the ants to find food and survive. According to the theory, a novel optimization algorithm, called CAS algorithm, was presented.

In the CAS algorithm, the chaotic system θ_{id+1} = θ_{id}+ exp(μ(1−θ_{id})) [28] was introduced into the heuristic equation of the CAS algorithm for obtaining the chaotic search initially. The adjustment of the chaotic behavior of individual ant is achieved by the introduction of a successively decrementation of organization variable μ_i and leads the individual to move to the new site acquired with the best fitness value eventually. \((p_{id}−θ_{id})\) is introduced to achieve the information exchange of individuals and the movements to new site taken on the best fitness value. \(p_{id}\) is selected based on the fitness theory which is very widely developed in optimization theory such as genetic algorithm and tabu search, and so on. \(θ_{id}\) is the state of the dth dimension of ant \(i\).

The CAS algorithm is a kind of iterative optimization algorithm, which is firstly employed in the optimization of sequential space. In the sequential space coordinates, the mathematic description [17] of the CAS algorithm as follows:

\[
\begin{align*}
\mu_i(n) &= \mu_i(n-1)(1+r_i) \\
\theta_{id}(n) &= \theta_{id}(n-1) + \frac{7.5}{\psi_{id}} \times V_i \\
e^{(1-e^{-2\psi_{id}(n+\delta)}}(p_{id}(n-1)−\theta_{id}(n-1))&\end{align*}
\]

where \(i=1,2,...,N\), \(N\) is the size of the ant swarm; \(d=1,2,...,L\), \(L\) is the dimension of the optimization space; \(n\) means the current iteration, and \(n-1\) is the previous iteration; \(\mu_i\) is the current state of the \(i\)th ant’s organization variable, \(\mu_i(0) = 0.999\); \(r_i\) is termed by us as the organization factor of ant \(i\); \(\psi_{id}\) determines the selection of the search range of the \(d\)th element of variable in the search space; \(V_i\) determines the search region of ant \(i\) and offers the advantage that ants could search diverse regions of the problem space. The value of \(V_i\) should be suitably selected according to concrete optimization problems; \(a\) is a sufficiently large positive constant and can be selected as \(a = 200\); \(\delta(0 \leq \delta \leq 2/3)\) is a constant; \(p_{id}(n-1)\) is the best position found by the \(i\)th ant and its neighbors within \(n-1\) steps; \(\theta_{id}\) is the current state of the \(d\)th dimension of ant \(i\), \(\theta_{id}(0) = (7.5/\psi_{id})(1-V_i)R\), where \(R\) is a uniformly distributed random number in \([0,1]\).

\(r_i\) and \(\psi_{id}\) are two important parameters. \(r_i\) is the organization factor of ant \(i\), which affects the convergence speed of the CAS algorithm directly. If \(r_i\) is very large, the iteration step of “chaotic” search is small then the system converges quickly and the desired optima or near-optima cannot be achieved. If \(r_i\) is very small, the iteration step of “chaotic” search is large then the system converges slowly and the runtime will be longer. Since small changes are desired as iteration step evolves, the value of \(r_i\) is chosen typically as \(0 < r_i \leq 0.5\). The format of \(r_i\) can be designed according to concrete problems and runtime. Each ant could have different \(r_i\), such as \(r_i = 0.3 + 0.02 \cdot \text{rand}(1)\). \(\psi_{id}\) affects the search ranges of the CAS algorithm. If the interval of the search is \([-\theta_{id}, \theta_{id}]\), then we can obtain an approximate formula \(\frac{\theta_{id}−\theta_{id}}{2} = 7.5\).

In principle, a neighborhood can be any ordered finite set. These neighbors are not necessarily individuals who are near them in the parameter space, but rather ones that are near them in a topological space. In fact the CAS algorithm does not impose any limitation on the definition of the distance between two ants. In order to simulate the behaviors of ants, we use the Euclidian distance. Supposing there are two ants whose positions are \((\theta_{i1},...,\theta_{id})\) and \((\theta_{j1},...,\theta_{jd})\), respectively, where \(i,j=1,2,...,N\) (where, \(N\) is the size of ant swarm) and \(i \neq j\), the distance between the two ants is \(\frac{1}{2}(\theta_{i1}−\theta_{j1})^2 +...+(\theta_{id}−\theta_{jd})^2\).

In the CAS algorithm, the neighbor selection can be defined in two ways. The first is the nearest fixed number of neighbors. The nearest \(m\) ants are selected as the neighbors of single ant. The second way is to consider the situation in which the number of neighbors increasing with iterative steps. This is due to the influence of self-organization behavior of ant \(i\). The impact of \(\theta_{i1}\) on the organization will become stronger than before and the neighbor of the ant will increase. That is to say, the number of nearest neighbor is dynamically changed as time evolves or iterative steps increase. The number \(q\) of
single ant is defined to increase for every $T$ iterative steps.

B. Parameters Optimization of LS-SVM Based on CAS Algorithm

As stated before, the CAS algorithm has powerful global search ability to find exact or approximate solutions for optimization and search problems. Thus, a parameters selection approach using the CAS algorithm for LS-SVM is presented in this paper. There are two key factors to determine the optimized parameters using the CAS algorithm: one is how to represent the parameters as the ant’s position, namely how to encode. Another is how to define the fitness function which evaluates the goodness of an ant. These two key factors are given as follows:

- **Encoding parameters:** the optimized parameters for LS-SVM include kernel parameter and regularization parameter. In solving parameters selection by the CAS algorithm, each ant is requested to represent a potential solution, namely parameters combination. So let us denote an $m$-parameters combination as a vector of dimension $m$. For example, if Gauss radial basis function (RBF) is chosen as a kernel function, we denote the vector as $\mathbf{v} = (\sigma, \gamma)$.

- **Fitness function:** the fitness function is a generalization measure. There are some different descriptions for the generalization measure. Therefore, the corresponding fitness can be determined. The fitness of an ant is evaluated by the mean square error (MSE) index, which is defined as the error between the function estimation of LS-SVM and the reference model. It can be expressed by

$$\frac{1}{N} \sum_{i=1}^{N} (y - f(x))^2$$

where $N$ denotes the number of training data, $y$ is the reference model, and $f(x)$ is the function estimation of LS-SVM.

In the CAS algorithm one aims at minimizing the MSE through choosing the optimal parameters combination, that is

$$\min f(z_1, \cdots, z_t) = \min \text{MSE}$$

subject to the equality constraints

$$g_i \leq z_i \leq h_i, i = 1,2$$

where the optimization variables are $\gamma$ and $\sigma$ respectively, $[g_i, h_i]$ denotes the value range for each variable, which is different with different reference model and training data.

The flowchart of the CAS-based parameters selection algorithm for the LS-SVM is shown in Fig. 1.

**IV. Simulation Research**

Experiment of a typical test function estimation is performed to evaluate the performance of the proposed parameters selection method. All experiments are performed on a PC with Pentium IV 2.93GHz processor, 512MB of main memory and the Matlab 6.5 simulation software.

Given one-dimensional Sinc function

$$f(x) = \text{sinc}(x) + \nu, x \in [-3,3]$$

where $\nu$ is the Gaussian noise with zero mean and standard deviation 0.1. We select 100 pairs of data as the train set from the input variable range. One aims at minimizing the MSE via the CAS algorithm to select the optimal kernel parameter $\sigma$ of Gauss RBF kernel function and regularization item $\gamma$. The searching ranges are set as follows: $\gamma \in [0.30], \sigma \in [0.5]$. The CAS algorithm parameters are chosen as follows: $N = 20$, the maximum number of iterations is 200, $\delta = 2/3$, $a = 200$, $r_i = 0.05 + 0.02 \times \text{rand}()$, $\psi_{11} = 0.25$, $\psi_{12} = 1.5$. In simulation, the first way is used to select neighbours of single ant. The researching results of parameters are $\gamma = 7.7379$ and $\sigma = 0.8851$, respectively.

The training result for LS-SVM via the above parameters is shown in Fig. 2. It can be seen from Fig.2 that LS-SVM realizes very good function approximation, so the CAS algorithm successfully realizes the parameters optimization selection for the test function.
In order to explain the effectiveness of this method, we adopt the genetic algorithm (GA, crossover rate is 0.8, mutation rate is 0.2%, population size is 30, the maximum number of iterations is 200) and particles swarm optimization algorithm (PSO, the population size and maximum number of iterations is the same as GA) to carry out many times’ experiments. The model of LS-SVM is tested with the testing set about 50 data produced by randomly initialized, the average results is recorded in Table 1. Table 1 shows the model testing MSE of this paper method is the minimum.

V. APPLICATION OF LS-SVM BASED ON THE CAS ALGORITHM

The inverted pendulum artificially created is a complex nonlinear system in order to deeply research the control for the nonlinear, high order and under-actuated system. Characterized as a typical nonlinear, high order, unstable and under-actuated system, it is very difficult to give a precise mathematical model. Therefore, the model identification research for the inverted pendulum system is very important.

The GPIP2003 single planar inverted pendulum is considered as a plant in the paper, whose inverse model is identified by LS-SVM. We adopt the example provided by the inverted pendulum toolbox, where the pendulum is displaced from lower position to the upper. After the pendulum reaches the upper position, one applies the disturbance by plucking the pendulum. The experiment data from the process overcoming the disturbance to the stabilization is sampled to the workspace of Matlab environment by the communication interface, which is stored as the text document by the command “save~”.

The data includes seven items such as the sampling period, the control variable, angle of the pendulum, position of the cart, angular rate of the pendulum, velocity of the cart and displacement of the objective. Angle of the pendulum, angular rate of the pendulum, position of the cart, velocity of the cart and the control variable are selected as multi-input and single-output model for LS-SVM. 100 pairs data from the input variable are chosen as the training sample set, in which 40 pairs data are selected as the testing sample set. The minimum MSE error as the fitness function, we utilize the CAS algorithm to carry the optimization selection for the regularization item $\gamma$ and kernel parameter $\sigma$. The researching results of parameters are $\gamma = 7.7379$ and $\sigma = 0.8851$, and the testing error is 0.0022. The estimation for the inverse model is achieved using the above result. The simulation result is shown in Fig. 3. Fig. 3 shows that the estimation value approaches to the real sampling value. simulation results show the LS-SVM model has good generalization performance and stronger robust performance after optimized by the CAS algorithm.

VI. CONCLUSION

Appropriate parameters are very crucial to least-squares support vector machines (LS-SVM) learning results and generalization ability. This paper presents a novel parameter selection method for LS-SVM is presented based on chaotic ant swarm (CAS) algorithm. The selection problem of LS-SVM parameters is considered as a swarm intelligence optimization problem, and a CAS optimization algorithm is employed to search the optimal objective function. CAS algorithm is global search method and it need not to consider LS-SVM dimensionality and complexity. Simulation and experiment results show that the proposed method is an effective approach for parameter optimization.
ACKNOWLEDGMENT

The authors are grateful to the anonymous referees for their valuable remarks and helpful suggestions, which have significantly improved the paper. This work was supported in part by a grant from the support of Key Project of Chinese National Programs for Fundamental Research and Development (973 Program) (2007CB7140 06), National Nature Science Foundation of China (61074020) and the Fundamental Research Funds for the Central Universities (DC10040101).

REFERENCES


Chunli Xie received his B.Sc. and M.Sc. degrees from Fushun Petroleum Institute and Liaoning Shihua University, Fushun, China, in 1993 and 2003, respectively. He is currently working toward the Ph.D degree with Dalian University of Technology, Dalian, China.

His research interests include adaptive control, robust control, machine learning, nonlinear systems, artificial intelligence and application.

Cheng Shao was born in Shenyang, P. R. China, on June 7, 1958. Shao received his B.Sc. degree from Liaoning University, Shenyang, China, in 1982. Then he received the M.Sc and Ph.D. degrees from Northeastern University, Shenyang, China, in 1987 and 1992. He is currently a full-time Professor and Ph.D. Advisor with School of Electronic and Information Engineering, Dalian
University of Technology, China. His research interest covers complex system modeling and intelligence control.

**Dandan Zhao** was born in Fuxin, P. R. China, on March 4, 1975. Zhao received her B.Sc. and M.Sc. degrees from Fushun Petroleum Institute and Liaoning Shihua University, Fushun, China, in 1997 and 2003, respectively. She is currently a full-time lecturer of School of Computer Science and Engineering, Dalian Nationalities University, Dalian, China. Her research interests include electronic commerce, semantic network, swarm intelligent and information processing.