Journal of Theoretical and Applied Information Technology
20th February 2013. Vol. 48 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

IDENTIFICATION OF NONLINEAR SYSTEM VIA SVR OPTIMIZED BY PARTICLE SWARM ALGORITHM

[1,2] XIANFANG WANG, [1] JIALE DONG, [1] YUANYUAN ZHANG, [3] ZHIYONG DU

[1] College of Computer & Information Engineering, Henan Normal University, Xinxiang 453007, China

[2] Henan Province Colleges and Universities Engineering Technology Research Center for Computing Intelligence & Data Mining, Xinxiang 453007, Henan, China

[3] Henan Mechanical and Electrical Engineering College, Xinxiang 453002, Henan, China

E-mail: xfwang11@yahoo.com.cn

ABSTRACT

Given the influence of the selection of regression parameters on the accuracy of SVR model and its ability of learning and generalization, this article adopts the particle swarm optimization algorithm to build the SVR model and applies it to the modeling of nonlinear system identification. Through the simulation experiments, it is found that this model is more accurate in identification and has a stronger ability of learning and generalization compared with GA. In addition, it demonstrates that the application in nonlinear system identification based on PSO-SVR algorithm could be considerably effective.

Keywords: *Particle Swarm Optimization (PSO), Support Vector Regression (SVR), Nonlinear System*

1. INTRODUCTION

With the rapid development of computer technology and control theory, system identification has become a subject of tremendous importance, which has been widely used in daily life and industrial production. It can be classified into linear system identification and nonlinear system identification. The theory system of linear system identification has become mature gradually. However, the nonlinear system identification field still has a lot of room for improvement because it's difficult to establish an accurate model considering the diversity and complexity in nonlinear system. As in practice most of the systems are nonlinear, the nonlinear system identification will be an important aspect for further research in the area of system identification. The theory, mainly based on neural network, is an effective tool to solve the problem concerning nonlinear system identification. But it is not flawless with problems such as over-fitting, local extremum, slow convergence rate, strong dependence on the quantity and quality of data.

SVM (Support Vector Machine) [1], a new machine learning algorithm based on the statistical learning theory, can get the global optimum solution without local extremum by using the structural risk minimization principle, which has distinct advantages in solving such practical problems as nonlinearity, small sample, and high dimension [2]. SVR (Support Vector Regression) is a regression algorithm established on the basis of SVM, which is applied in functional regression.

As the selection of regression parameters (ε, C, γ) has an enormous influence on the accuracy of SVR model and its learning generalization ability in the estimation of nonlinear support vector regression, it's necessary to optimize the parameters. GA (Genetic Algorithm) can be applied to the parameter optimization, but, due to its computational complexity it's not efficient enough in searching the optimal solution. PSO (Particle Swarm Optimization) which has stronger global searching ability and faster convergence speed than GA can realize the optimization of multiple parameters at the same time allowing the model to achieve better regression effect. Therefore, this article uses PSO to get the optimal parameters and model SVR, which is then applied to the nonlinear system identification by MATLAB simulating experiment.

2. ALGORITHM THEORY

2.1 Particle Swarm Optimization Algorithm

PSO is a swarm intelligence optimization algorithm first proposed by Kennedy and Eberhart in 1995 [3]. PSO algorithm establishes a simple velocity and displacement model to realize the optimization in the solution space without adjusting

Journal of Theoretical and Applied Information Technology
20th February 2013. Vol. 48 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

parameters. So the algorithm is easy to achieve with its faster convergence speed and has some advantages compared with other optimization algorithms. The basic idea of PSO algorithm is this: A group of particles are initialized in the entire solution space, and each of them, measured by velocity, position, and fitness value, may be an optimal solution for the problem. Then these particles adjust and update their own position dynamically according to the mobile experience of themselves and other particles around them. Each time, the particle would update the velocity and position of individual extremum (pbest) and group extremum (gbest) by comparing the fitness value of new particle with the fitness value of individual extremum and group extremum. The formulas are as follows:

$$v_i^{k+1} = w \times v_i^k + c_1 \times rand_1(pbest_i^k - x_i^k) + c_2 \times rand_2(gbest_i^k - x_i^k) \quad (1)$$

$$x_i^{k+1} = x_i^k + v_i^{k+1} \quad (2)$$

where k denotes the current iteration number of the particle, v_i^k, v_i^{k+1} are the current particle speed and the speed of next generation, x_i^k, x_i^{k+1} are the current particle position and the position of next generation, w which determines the impact of historical speed on current speed is the inertia weight, non-negative constants c_1 and c_2 denote learning factors, $rand_1, rand_2$ are the random numbers between 0 and 1, and $pbest_i^k, gbest_i^k$ are the individual extremum and the global extremum of the current particle [4].

2.2 Support Vector Regression

SVM, a machine learning algorithm based on statistical learning theory, was initially proposed for classification of problems by Vapnik et al. On the basis of SVM, SVR, which introduces loss function, is applied in the regression learning [5].

Firstly, the linear regression is discussed. A linear function $f(x) = w\ x + b$ is used to fit the training sample set $\{(x_i, y_i)\}, i = 1, 2, \ldots, l$, where w is the weight vector, b is the bias, x_i denotes input vector, and y_i is the output value of x_i. Slack variables ξ_i and ξ_i^* are introduced due to the error in fitting, and thus, the modeling problem is transformed into the optimization problem:

$$\min_{w,b,\xi,\xi^*} = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*) \quad (3)$$

$$s.t. \begin{cases} y_i - (w\ x_i + b) \le \varepsilon + \xi_i \\ (w\ x_i + b) - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \quad (i = 1, 2, \ldots, l) \\ C > 0 \end{cases} \quad (4)$$

where C denotes the penalty coefficient, ε is the insensitive loss function. When the difference between $f(x_i)$ and y_i is less than ε, the error is supposed to be zero, namely, no loss. Otherwise, the error is $|f(x_i) - y_i| - \varepsilon$.

To solve the problem of mathematical optimization more easily which is a convex quadratic programming problem Lagrange function and duality principle are used, then we can get its dual form as follows:

$$\max : L(\alpha, \alpha^*) = -\varepsilon\sum_{i=1}^{l}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{l} y_i(\alpha_i^* - \alpha_i) - \frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)(\alpha_i\ \alpha_j) \quad (5)$$

$$s.t. \begin{cases} \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0, & i = 1, 2, \ldots, l \\ 0 \le \alpha_i, \alpha_i^* \le C \end{cases} \quad (6)$$

By solving Lagrange multipliers α^* and α the following function to be estimated is obtained:

$$f(x) = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)(x_i\ x) + b \quad (7)$$

According to Equation (7), the linear regression function can be got:

$$f(x) = w\ x + b = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)(x_i\ x) + b \quad (8)$$

Next, the nonlinear regression is discussed. The nonlinear transformation is adopted to map the data to high dimensional space, thus translating it into the problem concerning nonlinear regression. The kernel function $K(x_i, x_j)$ is introduced here to calculate the inner product $\psi(x_i)\ \psi(x_j)$ in high dimensional feature space, and the nonlinear regression function is as follow:

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$f(x)=\sum_{i,j=1}^{l}(\alpha_i^*-\alpha_i)K(\alpha_i,\alpha_j)+b \qquad (9)$$

There are many common kernel functions, such as radial basis function (RBF), polynomial kernel function, and linear kernel function etc. According to researches and experiments, the results from using RBF are desirable in most cases. That's why Gaussian RBF $K(x_i,x_j)=\exp\frac{-|x_i-x_j|}{2\gamma^2}$ is chosen as kernel function in this paper.

3. SVR BASED ON PSO ALGORITHM

In the estimation of nonlinear support vector regression we are mainly concerned with the optimization of insensitive loss function ε, penalty coefficient C, and γ in kernel function $K(x_i,x_j)$, which are decisive to the SVR model in generalization ability and its learning accuracy [6]. Among the three parameters, ε affects the model accuracy: the smaller the ε is, the more support vectors we have and the more accurate the model is likely to be; C has a great influence on the generalization ability of the model: with the rise of C, the data's fitting degree tends to increase, but the generalization ability decreases; γ also concerns the learning accuracy of the model.

In order to find the optimal parameter combination of SVR model, PSO algorithm is used to optimize the three-dimensional parameter (ε,C,γ) [7]. As the velocity and position of each particle are determined by three-dimensional parameter (ε,C,γ), mean square error (MSE) which can reflect the performance of SVR regression is chosen as the fitness function Fit [8], that is:

$$Fit=MSE=\sqrt{\sum_{i=1}^{l}\frac{(y_i^*-y_i)^2}{l}} \qquad (10)$$

where l denotes the total number of samples, y_i is the actual value of the ith sample, and y_i^* is the corresponding output value of SVR model of the ith sample.

The detailed steps of SVR in parameter selection based on PSO are as follows [9]:

① The input vector and output vector of SVR need to be determined.

② PSO algorithm is adopted to find the parameters (ε,C,γ) of SVR model: Firstly, initialize the velocity and position of each particle, set the algorithm's iteration number and determine the population size. Secondly, calculate the fitness value of each particle, then search for individual extremum and group extremum on the basis of the fitness value of each initial particle. Thirdly, update the velocity and position of each particle according to (1) and (2), and renew individual extremum and global extremum based on the fitness values of particles in the new population. Finally, if the termination condition that the predetermined fitness value or the maximum iteration number can be reached is satisfied, the optimization will end, otherwise the calculation of the particles' fitness values is involved.

③ Based on the optimal parameter combination obtained from the above steps, the SVR model is established.

4. SIMULATION EXPERIMENT

4.1 Simulation Object

In order to verify the effectiveness of the application in nonlinear system based on SVR which is optimized by particle swarm algorithm, the SISO nonlinear system from the reference [10] is cited in this thesis:

$$y(k+1)=\frac{1.5y(k)y(k-1)}{1+y^2(k)+y^2(k-1)}-0.35\sin[y(k)+y(k-1)]+1.2u(k) \qquad (11)$$

4.2 The Selection of Parameters

First of all, PSO is adopted to optimize the parameters after the operating parameters of the algorithm are set, where the particle number is 20, the iteration number is 100, and both c_1 and c_2 are 2.

The actual value y_i of the model and the output value y_i^* of SVR model need to be plugged into the formula (10) to work out its MSE constantly. Then we can obtain the minimum MSE 0.003419 when the iteration reaches 39, and consequently get the optimal parameter combination $(0.001,600,3)$. The fitness curve of searching parameters with PSO is shown in Figure 1. To verify the result of PSO, algorithm GA is used to optimize the parameters, thus getting the Figure 2 which is the fitness curve of GA.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

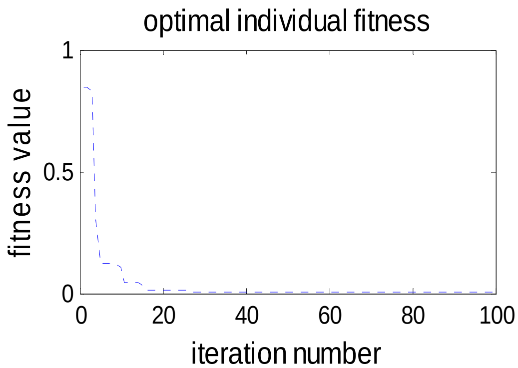

Figure 1: The fitness curve of PSO

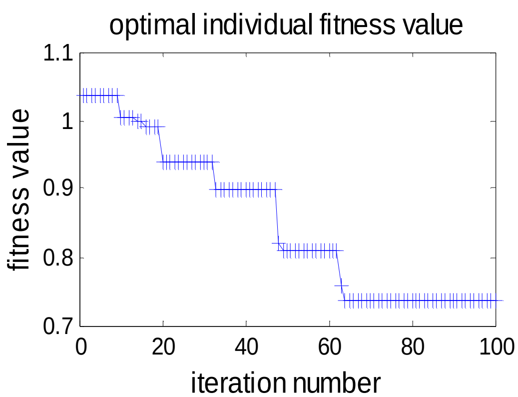

Figure 2: The fitness curve of GA

Table 1: Comparison between PSO and GA

Iteration number	MSE	
	PSO	GA
1	0.842	1.036798
10	0.1102	1.005336
20	0.01194	0.939499
30	0.003858	0.939499
39	0.003419	0.899560
60	0.003419	0.808606
100	0.003419	0.735759

Through comparison, it can be seen that both GA and PSO optimize the parameters iteratively, but the MSE of PSO is always lower than that of GA when their iterations are same. And also the optimum solution with PSO has appeared when the iteration number is 39, however, the optimal solution with GA hasn't surfaced until the iteration number is 100, which indicates that PSO has better convergence and takes less time than GA. And that is why PSO algorithm is introduced to optimize parameters.

4.3 Nonlinear System Identification Based on PSO-SVR

We need to plug the above optimal parameters into the given model and set the amplitude of white noise signal at 2. Figure 3 and Figure 4 are the output graph and the error image of the model.

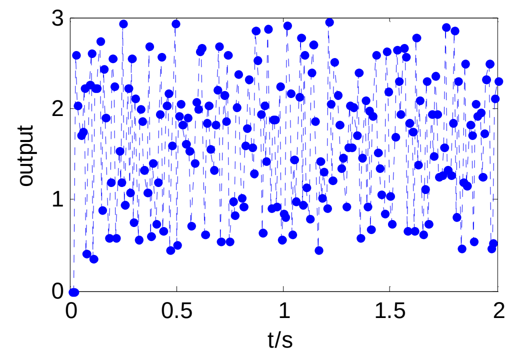

Figure 3: The Model Output In White Noise Signal

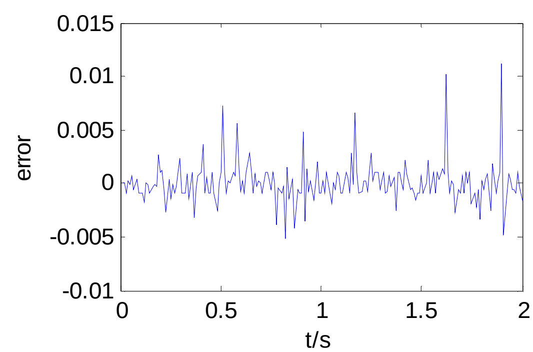

Figure 4: The Model Error In White Noise Signal

Figure 4 shows that the error of the model is kept within 10^{-3} orders of magnitude, which indicates that the model has a higher accuracy.

The random signal (amplitude 0.8), sinusoidal signal $(0.4\sin(2\pi t)+0.4)$ and square wave signal $(0.4 sign[\sin(2\pi t)]+0.4)$ are used to verify the generalization ability of the mode, and the results are shown in Figure 5 - Figure7.

Journal of Theoretical and Applied Information Technology
20th February 2013. Vol. 48 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

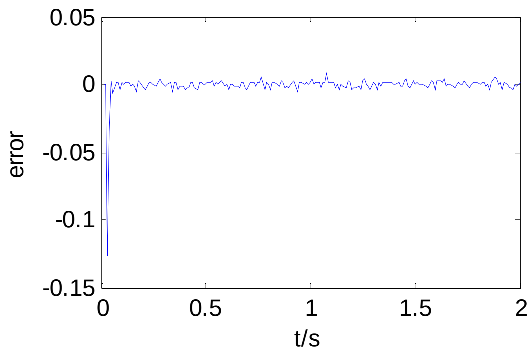

Figure 5: The Error Checking In Random Signal

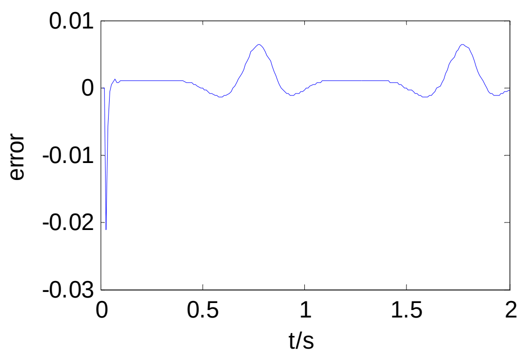

Figure 6: The Error Checking In Sinusoidal Signal

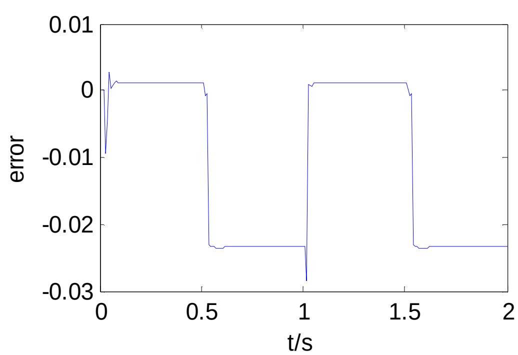

Figure 7: The Error Checking In Square Wave Signal

As seen in Figure 5 – Figure 7 the error remains less than 10^{-2} orders of magnitude under different calibrating signals, which shows that the model also has a better generalization ability.

5. CONCLUSION

As the optimized selection of parameters (ε, C, γ) exerts a great influence on the regression accuracy of SVR model and its learning and generalization ability in the estimation of nonlinear support vector regression, it's necessary to optimize the parameters. Therefore, PSO is introduced to obtain the optimal parameters and model SVR in this paper, which is then applied to the modeling of nonlinear system identification. The simulation results show that PSO is better in terms of convergence and more efficient in optimization compared with GA, which makes the model get higher identification accuracy and stronger ability of learning and generalization. Though the model takes less time compared to GA when the parameters of SVR are optimized, the total runtime of program is still rather long, which means the model has yet to be improved.

ACKNOWLEDGEMENTS

This work is supported by the National Natural Science Foundation of China (No.61173071), the Science and Technology Research Project of Henan Province(No.122102210079), the Foundation and Frontier Technology Research Programs of Henan Province(No.112400430087), the Innovation Talent Support Program of Henan Province Universities (No.2012HASTIT011), the Doctoral Started Project of Henan Normal University(No.1039).

REFERENCES:

[1] Vapnik V, "The nature of statistical learning theory", *Wiley-Inter Science*, New York, 1998.

[2] Ding Shifei, Qi Bingjuan, Tan Hongyan, "An overview on theory and algorithm of support vector machines", *Journal of University of Electronic Science and Technology of China*, Vol. 40, No. 1, 2011, pp. 2-8.

[3] KENNEDY J, EBERHART R C, "A discrete binary version of the particle swarm algorithm", *IEEE Service Center, NJ*, 1995, pp. 4104-4109.

[4] Li Aiguo, Qin Zheng, Bao Fumin, He Shengping, "Particle swarm optimization algorithms", *Computer Engineering and Applications*, Vol. 21, 2002, pp. 1-3.

[5] Desai K,Badhe Y,Tambe S S.et al, "Soft-sensor development for fed-batch bioreactors using support vector regression", *Biochemical Engineering Journal*, Vol. 27, No. 3, 2006, pp. 225-239.

[6] Ustun B,Melssen W J, "Determination of optimal support vector regression parameters by genetic algorithms and simplex optimization", *Analytical Chimica Acta*, Vol. 544, 2005, pp. 292-305.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[7] Hou Yongqiang, Wang Zengbao, "Active disturbance rejection controller parameters optimization based on particle swarm optimized", *Computer & Digital Engineering*, Vol. 40, No. 2, 2012, pp. 26-28.

[8] Mao Zhiliang, Liu Chunbo, Pan Feng, "Parameter selection and application of SVM with mixture kernels based on IPSO", *Journal of Jiangnan University (Natural Science Edition)*, Vol. 8, No. 6, 2009, pp. 632-634.

[9] Chen Shu, Xu Baoguo, Wang Haixia, Wu Xiaopeng, "Study of fermentation process based on PSO-SVR", *Computer Engineering and Applications*, Vol. 43, No. 19, 2007, pp. 214-216.

[10] Zhang Haoran, Han Zhengzhi, Li Changgang, "Support vector machine based nonlinear systems identification", *Journal of System Simulation*, Vol. 15, No. 1, 2003, pp. 119-125.