Available online at www.sciencedirect.com

SciVerse ScienceDirect

Physics Procedia 25 (2012) 588 – 594

Physics Procedia

2012 International Conference on Solid State Devices and Materials Science

Chaos Optimization Method of SVM Parameters Selection for Chaotic Time Series Forecasting

Hu Yuxia[1], Zhang Hongtao[2]

[1]College of Electric Engineering , Zhengzhou University, Zhengzhou, China

[2]Institute of Electric power , North China Institute of Water Conservancy and Hydroelectric Power, Zhengzhou, China

Abstract

For support vector regression (SVR), the setting of key parameters is very important, which determines the regression accuracy and generalization performance of SVR model. In this paper, an optimal selection approach for SVR parameters was put forward based on mutative scale optimization algorithm(MSCOA), the key parameters C and ε of SVM and the radial basis kernel parameter g were optimized within the global scopes. The support vector regression model was established for chaotic time series prediction by using the **optimum parameters**. The time series of Lorenz system was used to testify the effectiveness of the model. The root mean square error of prediction reached $RMSE = 3.0335 \times 10^{-3}$. Simulation results show that the optimal selection approach based on MSCOA is an effective approach and the MSCOA-SVR model has a good performance for chaotic time series forecasting.

Keywords: support vector machine; Mutative scale chaos optimization algorithm; chaotic time series prediction; phase space reconstruction, parameter selection.

1. Introduction

With the development of chaos theory and its application, the analysis and forecasting of chaotic time series has become a very important research direction in signal processing field in recent years. The prediction of chaotic time series has been used to stock finance, electricity load forecasting, geological environment, weather forecasting [1-4].

Support Vector Machine (SVM) is new machine learning based on statistical theory[5]. It can solve small-sample, non-linear and high dimension problems by using structural risk minimization (SRM) instead of empirical risk minimization (ERM). Its stronger generalization ability and very good application potentiality has been shown in classification and regression. The key parameters of support vector machine are very important, the accuracy of classification or regression is determined by a group of appropriate parameters. In recent years many researches on model selection have been done including grid search[6], mutative scale chaos optimization algorithm[7], genetic Algorithm[8], etc.

 doi:10.1016/j.phpro.2012.03.130

The chaos optimization algorithm(COA) is an efficient and convenient way for global optimization, and to improve the search efficiency and accuracy, mutative scale chaos optimization is employed, which can reduce the search ranges during the search process. The one-dimensional Logistic map is usually employed in chaos optimization algorithm[9-10].

In this paper, the mutative scale optimization algorithm was used to select parameters of SVM for chaotic time series forecasting, and the key parameters C 、 ε and the radial basis kernel function parameter g were optimized and trained by using COA based on Logistic map. Considered the selection of key parameters of SVM an optimization grouping problem, and the objective function was established for this problem. The time series of Lorenz system was used to testify the effectiveness of the models, and simulation results show that the optimal selection approach of SVM key parameters based on COA is available and the SVM model has a good performance for chaotic time series forecasting.

2. SVM regression (SVR) theory

Prediction of chaotic time series can be attributed to support vector machine regression problems. In SVR, the basic idea is to map the data into a higher dimensional feature space via a nonlinear mapping $\Phi(\bullet)$ and then to do linear regression in the space. Therefore, regression approximation addresses the problem of estimating a function based on a given data set $\{\mathbf{x}_i, y_i;\ i = 1, 2, \cdots, N\}$ (where $\mathbf{x}_i \in R^n$ is the input vector and $y_i \in R$ is the desired value). SVM approximates the function with the form

$$f(\mathbf{x}) = \mathbf{w}^{\mathrm{T}} \bullet \Phi(\mathbf{x}) + b, \quad \Phi{:}R^n \to F, \mathbf{w} \in F \tag{1}$$

Where $\{\Phi_i(\mathbf{x})\}_{i=1}^{N}$ are the points in the features space, $\{\mathbf{w}_i\}_{k=1}^{N}$ and b are coefficients. They can be estimated by minimizing the regularized risk function:

$$R(C) = C\frac{1}{l}\sum_{i=1}^{N} L_{\varepsilon}(y_i, f(x_i)) + \frac{1}{2}\|\mathbf{w}\|^2 \tag{2}$$

Where $L_{\varepsilon}(y_i, f(x_i))$ is loss function measuring the approximation errors between expected output y_i and calculated output $f(\mathbf{x}_i)$, and C is regularization constant determining the tradeoff between the training error and the generalization performance. The second term $\frac{1}{2}\|\mathbf{w}\|^2$ is used as a measurement of function flatness. Introduction of relaxation factor ξ, ξ^* leads Eq. (2) to the following constrained function:

$$\begin{aligned} &\min\ J(\mathbf{w}, \xi, \xi^*) = \frac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w} + C\sum_{i=1}^{N}(\xi + \xi^*) \\ &\text{s.t.} \quad \begin{cases} y_i - \mathbf{w}^T \mathit{\Phi}(\mathbf{x}_i) - b \le \varepsilon + \xi_i \\ \mathbf{w}^T \mathit{\Phi}(\mathbf{x}_i) + b - y_i \le \varepsilon + \xi^*{}_i \\ \xi, \xi^* \ge 0 \end{cases} \end{aligned} \tag{3}$$

Finally, by introducing Lagrange multipliers and exploiting to the optimality constraints, the decision Eq. (1) has become the follow form:

$$f(\mathbf{x}) = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)K(\mathbf{x}_i, \mathbf{x}) + b \tag{4}$$

Where $\alpha_i, \alpha_i^*, i = 1, \cdots, N$ are Lagrange multipliers, satisfying the equalities $\alpha_i \times \alpha_i^* = 0$, $\alpha_i > 0$ and $\alpha_i^* > 0$. They are obtained by maximizing the dual formula of Eq. (3), which has the following form:

$$\begin{aligned} \max L(\boldsymbol{\alpha}, \boldsymbol{\alpha}^*) = & -\frac{1}{2}\sum_{i=1}^{N}(\alpha_i - \alpha_i^*)(\alpha_i - \alpha_i^*)K(\mathbf{x}_i, \mathbf{x}_l) \\ & -\varepsilon \sum_{i=1}^{N} y_i(\alpha_i - \alpha_i^*) + \sum_{i=1}^{N} y_i(\alpha_i - \alpha_i^*) \end{aligned} \tag{5}$$

$$\text{s.t.} \quad \begin{cases} \sum_{i=1}^{N}(\alpha_i - \alpha_i^*) = 0 \\ 0 \le \alpha_i, \alpha_i^* \le C, \quad i = 1, 2, \cdots, N \end{cases}$$

According to the nature of SVM regression, most of the α_i and α_i^* are zeros. Hence, the final formulation can be arrived by solving optimization problem mentioned above and the support vector regression has the following form:

$$f(\mathbf{x}) = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)K(\mathbf{x}, \mathbf{x}_i) + b \tag{6}$$

Although nonlinear function $\Phi(\bullet)$ is usually unknown, all computations related to $\Phi(\bullet)$ could be reduced to the form $K(\mathbf{x}_i, \mathbf{x}) = \Phi(\mathbf{x}_i)^{\mathrm{T}} \bullet \Phi(\mathbf{x})$. So what we need to do is selecting the appropriate kernel function. The advantages of RBF kernel function make it universally applied to SVM. The RBF kernel function can be applied to any sample by choosing right parameters.

3. Prediction model for chaotic time series based on SVM

A. Phase space reconstruction

Chaotic time series prediction is based on the theory of phase space reconstruction under Takens embedding theorem. A reconstructed phase space is a m-dimensional metric space into which a time series is embedded[11].

Given an observation chaotic time series $\{x(t)\}, t = 1, 2, \cdots, N$, selecting the embedding dimension m, the delay time τ, the phase space can be expressed as the following form:

$$\begin{aligned} X(t) = (x(t), x(t-\tau), \cdots x(t-(m-1)\tau), \\ X(t) \in R^m, (t = 1, 2, \cdots, M) \end{aligned} \tag{7}$$

Where $X(t)$ is a vector or a point in the construction phase space, $M = N - (m-1)\tau$ is the number of points in the reconstructed phase space. If the embedding dimension is large enough, the phase space is homeomorphic to the state space that generated the time series, that is, the phase space contains the same information as the original state space. There is a determinism map $f(\bullet)$ meeting the following equation:

$$x(t+T) = f(X(t)) \tag{8}$$

Local model is usually used to predict the next-step, in which, $T > 0$ is the step of forward prediction. $f(\bullet)$ is the predicting model for chaotic time series.

B) SVM predicting model for chaotic time series

Setting the $f(\bullet)$ of Eq. (8) as the support vector regression corresponding to Eq. (6) and $T=1$, the one-step prediction model for chaotic time series based on SVM has the following form:

$$\hat{x}(t+1)=\sum_{i=1}^{M}(\alpha_i-\alpha_i^*)\mathrm{K}(X(t),X(i))+b \tag{9}$$

where $\hat{x}(t+1)$ is the value of one-step forecasting, then the next points in phase space can be expressed as $\hat{X}(t+1)=(\hat{x}(t+1),x(t+1-\tau),\cdots x(t-(m-2)\tau))$ 。

4. COA for Parameters Optimization of SVM

A. The establishment of objective function

The absolute error is usually taken to measure the accuracy of the prediction model

$$e(i)=y(i)-\hat{y}(i) \tag{10}$$

Where $y(i)$ and $\hat{y}(i)$ were the actual values and predicted values of chaotic time series. The root mean square error (RMSE) is always used as integral performance index in prediction of chaotic time series.

$$RMSE=\sqrt{\sum_{1=1}^{N}(y(i)-\hat{y}(i))^2/N} \tag{11}$$

Where N was the number of test points.

In this paper, the root mean square error was used for the objective function.

$$J=\begin{cases}\min f(x_1,x_2,x_3)=\min(RMSE)\\ \text{s.t.}\quad x_i\in[a_i,b_i],\ \ i=1,2,3\end{cases} \tag{12}$$

Where x_1,x_2,x_3 corresponded to the parameters C , ε and g , $[a_i,b_i]\subset R,\ \ i=1,2,3$ are the value ranges of C , ε and g , and f is a real-valued continuous function.

B. The chaos optimization process

Chaos systems have special characters such as the ergodic property, stochastic property and sensitivity dependence on initial conditions of chaos.

One-dimensional Logistic map is usually applied to chaos optimization algorithm. The mathematical model of Logistic map is the following form:

$$\mathrm{Z}_{k+1}=\mu\mathrm{Z}_k(1-\mathrm{Z}_k),\ \ \mathrm{Z}_k\in[0,1],\ \ k=1,2,\cdots, \tag{13}$$

Where Z_k is the value of the variable Z at the kth iteration, Z_k in the interval [0, 1], μ is a so-called bifurcation parameter of the system ($0\le\mu\le4$). If $3.5699456\cdots\le\mu\le4$, Logistic mapping works in chaotic state, that is, no steady-state solution, the system is a full map interval of [0,1].

In this paper the chaos optimization algorithm was improved based on the literature [9] described as follows.

Step 1: Algorithm initialization. Given small different initial values $Z_i^0\subset(0,1), i=1,2,\cdots,m$ (which are not $0, 0.25, 0.5, 0.75, 1, (2+\sqrt{3})/4$). Generate the different chaotic variables Z_i^k , $k=1,2,\cdots,N$ by

Eq.(13), where N is the length of chaotic time series, k is a random integer in set $\{1,\cdots,N\}$. The initial solution $\mathbf{x}_0=(x_{0,1},x_{0,2},\cdots,x_{0,m})^{\mathrm{T}}$ is produced by the formula

$$x_i^0=a_i+(b_i-a_i)Z_i^k,\qquad i=1,2,\cdots,m \tag{14}$$

Where k is a random integer in set $\{1,\cdots,N\}$. Set $\mathbf{x}_0$ as the initial optimal solution x^* and $f(\mathbf{x}_0)$ as the initial smallest objective function f^*. In addition, a termination criterion is created: maximal iteration time M which is larger, maximal acceptable mean square error J_{end} or the number of times achieving the optimal solution A. Set K as the number of iterations that start from 1 and the times achieving the optimal solution $time=0$.

Step 2: Conversion of search ranges.

$$x_i^k=a_i+(b_i-a_i)Z_i^k,\qquad i=1,2,\cdots,m \tag{15}$$

Hence, the search ranges of *i*th chaos variables x_i^k will be changed from [−1, 1] to $[a_i,b_i]$.

Step 3: Coarse search with chaotic variables.

Set $x_i(K)=x_i^k$ and compute $f(x_i(K))$;

If $f(x_i(K))\le f^*$, then $f^*=f(x_i(K))$, $x^*=x_i(K)$, $time=time+1$;

Else $f(x_i(K))>f^*$, f^* and x^* is maintained.

If $time>A$, then goes to Step 4.

Step 4: If $K>M$ or $f^*<J_{\text{end}}$, this iteration is stopped; If $K<M$ and $f^*>J_{\text{end}}$, $K=K+1$, the iteration is to be continued and the search range scare modified

$$a_i'=x_i^*-\frac{1}{K+1}(b_i-a_i),\ b_i'=x_i^*+\frac{1}{K+1}(b_i-a_i) \tag{16}$$

To avoid a_i' and b_i' exceeding ranges $[a_i,b_i]$ [ai, bi], the updated ranges are restricted to their bounds:

If $a_i'<a_i$, Then $a_i'=a_i$; If $b_i'>b_i$, Then $b_i'=b_i$.Go back to Step 2 for next iteration.

*Step*5: Output the best solution $\mathbf{x}^*$ and the best value f^*.

C. the parameters setting

In Matlab environment, the LIBSVM developed by scholars Chih-Jen Lin was applied to realize chaotic time series prediction. Setting the ranges of the parameters: $C\in(2^{-5},2^{15})$, $\varepsilon\in(2^{-12},2^{0})$, $g\in(2^{-15},2^{5})$, It was evident that ranges of the parameters were large enough to meet most systems. The maximal iteration time $M=10000$; The maximal acceptable mean square error $J_{\text{end}}=0.001$ and the number of times achieving the optimal solution $A=10$.

5. Experiments

Lorenz attractor was taken to examine the SVM prediction model optimized by CSAA whose formula is following form:

$$\begin{cases} \dot{x} = a(y - x) \\ \dot{y} = -xz + rz - cy \\ \dot{z} = xy - bz \end{cases} \tag{17}$$

The system is chaotic when $a = 10$, $b = 8/3$, $c = 28$. And the differential equations were solved numerically using 4th order Runge-Kutta integration with a step size $\Delta t = 0.01$ and initial value $x(0) = 0$, $y(0) = 1$, $z(0) = 0$. In order to eliminate the influence of the initial value, the initial 1000 points were discarded, the phase space reconstruction of x chaotic time series was done with the embedding dimension m, the delay time τ. Then 1500 points were taken as training data, 1000 points were taken as test data.

Normalizing dates to [0 1] and training SVR with COA, the optimal values of C, ε, g were obtained, $C = 470.98, \varepsilon = 0.008186, g = 0.037806$, Then the prediction experiment was done with the SVR model optimized. $RMSE = 3.0335 \times 10^{-3}$, and the prediction and absolute error curves was shown in Figure 1 and Figure 2.

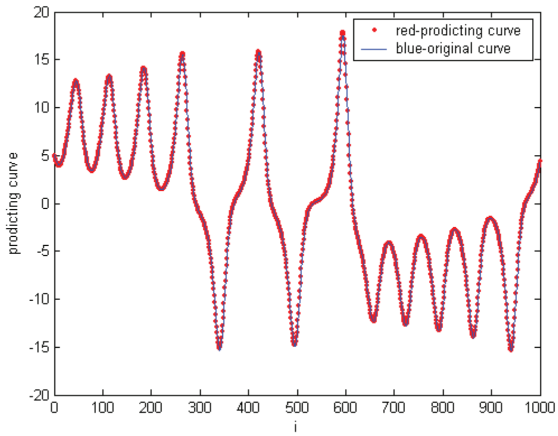

Fig 1. Predicted results and the original curve of Lorenz system

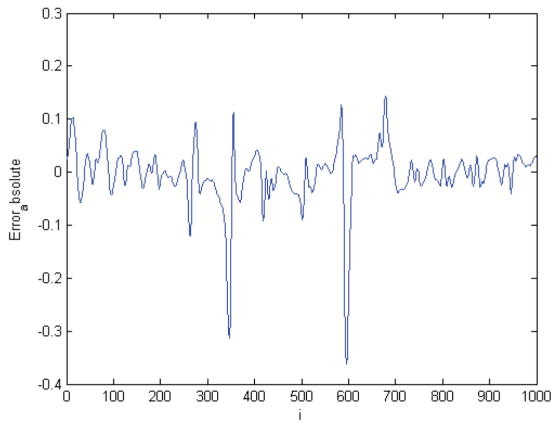

Fig 2. Absolute prediction error curve of Lorenz system

6. Conclusion

This article gave a simple introduction to the significance and the application of chaotic time series prediction, and the basic theory of support vector machine regression was introduced; the working principle of chaos optimization algorithm and mutative scale chaos optimization algorithm working process was explained detailed. The SVR model of chaotic time series was established by mutative scale chaos optimization algorithm, and the chaotic time series of Lorenz system was employed to examine the SVR model. Simulation results show the effectiveness of the algorithm.

Acknowledgements

This work was supported by National Natural Science Foundation of China (Grant No.30871449) and Education Department of Henan Province Natural Science Research projects (No .2008B510014).

References

[1] Q. Huang，W. Xie, “The prediction of short-term exchange rate based on the phase space reconstruction and kalman,” *Computer Applications and Software*, vol.25, no.8, pp79-80, August 2008.

[2] Z. Zhang, Y. Sun, “A new approach of STLF based on combination of phase space reconstruction theory and optimal recursive neural networks,” *Electric Power*, vol.37, no. 1, pp19-23, January 2004.

[3] Z. Shan, C. Lin, “Chaos Prediction of Wave Hydrodynamic Pressure Signals Based on Local Support Vectors Machine,” *Journal of System Simulation*, 2008, vol.20, no.23, pp6470-6472, December 2008.

[4] G. Wu, P. Zhou, “Prediction Based on Phrase Construction and Its Application in Weather Forecast,” *Chinese journal of nature*, vol. 21, no. 2, pp107-110, April 999.

[5] Cortes C, Vapnik V N, “Support Vector Networks,” Machine Learning, Vol.20, no.3, pp273-295, September 1995.

[6] C. Hsu, C. Chang , C. Lin, “A practical guide to support vector classification,” http://www.csie.ntu.edu.twcjlin/papem/guide/guide.pdf, 2003

[7] Q. Liu, P. Que, C. Fei, “Model selection for SVM using mutative scale chaos optimization algorithm,” *Journal of shanghai university*, vol. 10, no. 6, pp531-534, December 2006. 36(4), P101-103.

[8] C. Huang C. Wang, “A GA-based feature selection and parameters optimization,” *Expert Systems with Applications*, no. 31, pp231–240, November 2006.

[9] X. Yuan, Y. Wang, “Parameter selection of support vector machine for functionapproximation based on chaos optimization,” *Journal of Systems Engineering and Electronics*, Vol.19, no.1, pp191－197, January 2008.

[10] Q. Liu, P. Que, C. Fei, S. Song , “Model selection for SVM using mutative scale chaos optimization algorithm,” *Journal of Shanghai University*, Vol. 10, no. 6,pp531-534,June 2006.

[11] F. Takens, in: D. Rand, L.-S. Young (Eds.), Dynamical Systems and Turbulence, Warwick, 1980, Springer, Berlin, 1981, pp366.