Chaos, Solitons and Fractals 89 (2016) 153–159

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

Nonlinear Science, and Nonequilibrium and Complex Phenomena

journal homepage: www.elsevier.com/locate/chaos

Study on network traffic forecast model of SVR optimized by GAFSA

Yuan Liu*, RuiXue Wang

School of Digital Media, Jiangnan University, Wuxi, Jiangsu 214122, China

ARTICLE INFO

Article history:
Available online 22 December 2015

Keywords:
Network traffic forecast
Optimization of parameters
Support vector regression (SVR)
Global artificial fish swarm algorithm (GAFSA)
Self-similarity

ABSTRACT

There are some problems, such as low precision, on existing network traffic forecast model. In accordance with these problems, this paper proposed the network traffic forecast model of support vector regression (SVR) algorithm optimized by global artificial fish swarm algorithm (GAFSA). GAFSA constitutes an improvement of artificial fish swarm algorithm, which is a swarm intelligence optimization algorithm with a significant effect of optimization. The optimum training parameters used for SVR could be calculated by optimizing chosen parameters, which would make the forecast more accurate. With the optimum training parameters searched by GAFSA algorithm, a model of network traffic forecast, which greatly solved problems of great errors in SVR improved by others intelligent algorithms, could be built with the forecast result approaching stability and the increased forecast precision. The simulation shows that, compared with other models (e.g. GA-SVR, CPSO-SVR), the forecast results of GAFSA-SVR network traffic forecast model is more stable with the precision improved to more than 89%, which plays an important role on instructing network control behavior and analyzing security situation.

1. Introduction

A network traffic forecast model with high precision plays an important role on understanding the network behavior to occur, analyzing the network security situation and guiding the network security detection and control. At present, with the pace of broadband network, the traffic data is becoming huge. The changes of data gradually become complex and difficulties in management also increase. It has become one of the hot spot field to find a more accurate and stable network traffic forecast model.

The existing network traffic forecast model can be classified as two kinds: linear and non-linear ones. The typical linear models include auto regressive (AR) model [1] based on short-range correlation properties, auto regressive moving average (ARMA) model [2], auto regressive integrated moving average (ARIMA) model [3], and fractional auto regressive integrated moving average (FARIMA) model [4] based on long-range correlation properties, etc. The single linear model is simple and easy to implement, but it is difficult to describe and forecast the increasing complex network traffic characteristics. Meanwhile the typical non-linear models include grey model [5,6], wavelet forecast model [7], neural network forecast model [8–10], support vector regression (SVR) model [11], and integration or improvement of these models. SVR forecast model has strong ability of generalization which can find the global optimal solution [12], and avoid the "Curse of dimensionality". The result of forecast is also better than other non-linear models, and it has been widely used [13–15]. There are some research on the improvement of SVR forecast model mainly including the improvement of the SVR model and the improvement of the parameters selection of SVR model. In the forecast process of SVR model, the selection of parameters has a significant effect on the result of forecast. At present, the intelligent

* Corresponding author. Tel.: +86 13806184617.
E-mail address: lyuan1800@sina.com (Y. Liu).

http://dx.doi.org/10.1016/j.chaos.2015.10.019
0960-0779/© 2015 Published by Elsevier Ltd.

algorithms used in SVR parameter selection mainly include genetic algorithm(GA) and particle swarm optimization(PSO). The SVR forecast model optimized by GA (GA-SVR) [16,17] is complex to set parameters. The optimization method of this algorithm is determined by the probability, and the results are not stable. The SVR model optimized by PSO (PSO-SVR) [18] is easy to fall into local extremum. The SVR model optimized by chaos particle swarm optimization (CPSO-SVR) [19] improves the forecast precision of the PSO-SVR model, but the stability of this model still remains not high.

Artificial fish swarm algorithm (AFSA) [20] raised by Dr XiaoLei Li in 2002, is a modern heuristic search algorithm and make a significantly effect of optimization, with characteristics such as parallelism, simplicity, quickly jumping out of local extremum and so on. The model is strong robustness, good global convergence and initial value insensitive. Wang and Gao [21] proposed an improved algorithm of AFSA – global artificial fish swarm algorithm (global artificial fish swarm algorithm, GAFSA), adding the global information to the process of seeking optimization solution, so that it makes the artificial fish swarm algorithm works better on optimization and have stronger convergence.

Our contributions can be summarized as follows: improving the forecast model Of SVR algorithm based on the novel optimization strategy of global artificial fish swarm algorithm, using GAFSA algorithm to conduct parameter optimization, training the SRV algorithm with the optimum parameters searched by GAFSA algorithm, establishing the forecast model of GAFSA-SVR to forecast the network traffic.

2. Basic SVR regression forecast model

The basic idea of SVR algorithm is using a nonlinear mapping to map the data to a high-dimensional feature space, free from the complicated calculation of dot product in the high-dimensional space. The basic process is as follows:

Dataset $T=\{(x_1,x_2),(x_1,x_2),\ldots,(x_l,x_l)\}$, x_i is input value, y_i is the corresponding forecast values, $i=1,2,\ldots,l$. Set the regression function $f(x)=w\times\varphi(x)+b$, where w is autoregressive coefficient or weight vector, and b is error value. Through risk minimization regularization function training parameter w and b confirm the function is

$$\min \frac{1}{2}\|w\|^2+C\frac{1}{k}\sum_{i=1}^{k}\varepsilon(f(x_i)-y_i)$$
$$s.t.\quad \varepsilon(f(x_i)-y_i)=\begin{cases}|f(x_i)-y_i|-\varepsilon & |\omega\times\varphi(x)+b-y_i|\ge\varepsilon\\ 0 & |\omega\times\varphi(x)+b-y_i|<\varepsilon\end{cases}\quad (1)$$

Among those, $\varepsilon(\cdot)$ is the ε-non-sensitive loss function proposed by Vapnik et al., and ε is ε-intensive loss parameter. C is penalty factor to balance positive experience risk and confident range so that the risk is minimum.

Convert the problem seeking optimal hyperplane to

$$\min_{w,b,\xi_i,\xi_i^*}\frac{1}{2}\|w\|^2+c\sum_{i=1}^{k}\left(\xi_i+\xi_i^*\right)$$
$$s.t.\begin{cases}y_i-w\times\varphi(x)-b\le\varepsilon+\xi_i & \varepsilon_i\ge 0, i=1,2,\ldots,l\\ w\times\varphi(x)+b-y_i\le\varepsilon+\xi_i^* & \varepsilon_i^*\ge 0, i=1,2,\ldots,l\end{cases}\quad (2)$$

Among those, $w\in R^n$, $b\in R$; φ is the map which maps the input data from lower-dimensional non-linear regression problems to high-dimensional feature space transforming to linear problems; ε and ε^* are non-negative slack variable.

Introducing Lagrangian multiplier a_i and a_i^*, to convert (3) to dual problem:

$$\min_{a^{(*)}\in R^{2l}}\frac{1}{2}\sum_{i,j=1}^{l}\left(a_l^*-a_l\right)\left(a_j^*-a_j\right)K\left(x_i,x_j\right)$$
$$+\varepsilon\sum_{i=1}^{l}\left(a_i^*+a_i\right)-\sum_{i=1}^{l}\left(a_i^*-a_i\right)$$
$$s.t.\begin{cases}\sum_{i=1}^{l}\left(a_i^*-a_i\right)=0\\ 0\le a_i^{(*)}\le C, i=1,2,\ldots,l\end{cases}\quad (3)$$

Among those, $K(x_i,x_j)$ is kernel function. The resulting regression function $f(x)$ is

$$f(x)=\sum_{i=1}^{l}\left(a_i^*-a_i\right)(\varphi(X_i),\varphi(X))+b \quad (4)$$

3. SVR forecast model optimized by GAFSA algorithm

3.1. Related parameters in the SVR forecast model

The selection of insensitive factor ε, penalty factor C and the kernel function is an important factor affecting how to choose the parameters of the SVR forecast model. Frequently used kernel functions include linear kernel, polynomial kernel, RBF(radial basis function) kernel, sigmoid kernel, Fourier Series kernel, spline kernel, etc. RBF kernel is the most widely used kernel function among them, which can be applicable to any sample by choosing proper parameters. This thesis will use RBF kernel as the kernel function of SVR forecast model. Here's the form of RBF kernel function:

$$K(x,x_i)=\exp\left[\frac{-|x-x_i|^2}{\sigma^2}\right] \quad (5)$$

The parameter σ of RBF kernel implicitly determines the distribution of the data mapped to the new space. In summary, the forecast performance of SVR model is determined by RBF parameter σ, insensitive factor σ and penalty factor C.

3.2. Basic artificial fish swarm algorithm

Artificial fish swarm algorithm is a kind of swarm intelligence optimization based on animal's behaviors, which can be optimized in the searching field by simulating the foraging, herding, rear chasing and random behavior of fish.

The foraging behavior assumes the current state of an artificial fish i as X_i, its range of vision as *visual* and the moving step length as *step*. Then we randomly choose a state X_j in its visual range:

$$X_j=X_i+visual\times rand(\,) \quad (6)$$

If the solution of this state is superior to that of state X_i, then step forward in this direction:

$$X_i^{t+1}=X_i^t+\frac{X_j-X_i^t}{\|X_j-X_i^t\|}\times step\times rand(\,) \quad (7)$$