

Network Traffic Prediction Based on SVR Improved By Chaos Theory and Ant Colony Optimization

Yonglin Liang¹ and Lirong Qiu²

¹*College of Computer Science, Shaoguan University
Daxue Street, Zhenjiang District, Shaoguan*

²*School of Information Engineering, Minzu University of China
Zhongguancun Street, Haidian District, Beijing*

¹*liangyonglin@126.com, ²qiu_lirong@126.com*

Abstract

Network traffic prediction is one of the significant issues. The model for network traffic prediction should meet the following requirements. First, the model should be taken into consideration the characteristics of the network flow such as burstiness, long-range dependence, periodicity and self-similarity. To achieve this, we decompose the original flow in a multi-scale manner into a set of linear and stable representations, and introduce chaos theory to improve the diversity and search coverage. Second, the model should be efficient and accurate. To this end, we propose a prediction model based on SVR, and utilize Ant Colony Optimization (ACO) algorithm for parameter selection of SVR. Besides, we conduct experiments to evaluate the proposed model.

Keywords: *Network traffic prediction, SVR, Chaos theory, Ant Colony Optimization (ACO)*

1. Introduction

Network traffic prediction is one of the significant issues in many areas, such as traffic engineering, congestion control and network management. Research shows that patterns emerge even when the network flow suddenly changes, which provides feasibility to analyze and predict the network traffic. Once the network traffic is predicted in an early stage, efficient congestion control mechanism is designed, and resources are properly allocated and scheduled. Accordingly, congestion can be reduced or avoided, and the utilization of network resources can be improved. Therefore, network traffic prediction has always been an interesting problem for researchers.

Generally, an ideal model of network traffic prediction should satisfy the following requirements. First, the model should precisely describe the statistical characteristics related to network flow. Other than burstiness, long-range dependence and periodicity, there also exists self-similarity in high speed network. Therefore, signals can be decomposed based on the self-similarity in a multi-scale manner to reduce the complexity of the data, and transform the non-linear unstable data into linear and stable. Moreover, it is already proved by [1] that self-similarity is closely connected to chaos. Considering the nature of chaos, such as randomness, ergodicity and sensitivity to the initial conditions, we propose to introduce chaos theory to improve the diversity and search coverage after the decomposition of the unstable flow sequence.

Second, the model should be efficient, both accurate and fast. Existing models of network traffic prediction include ARMA model [2], BP neural network [3], grey model [4], and Support Vector Machine (SVM) [5], *etc.* However, although ARMA is simple and easy to implement, it is only suitable for linear prediction and the accuracy is not good enough [6]. BP neural network is based on the principle of minimizing empirical risks,

and has some problems such as local minimum and weak generalization, and therefore the prediction accuracy is limited [7]. The data samples for grey model are small and the modeling process is simple, but it cannot exhibit good accuracy with volatile data [8]. SVM can find the global best solution to deal with the small sample issue, non-linear and high-dimensional problems, but Support Vector Regression (SVR) is more suitable for trend prediction scenario based on SVM [9]. However, the parameter selection of SVR plays a decisive role in the learning accuracy and generalization ability of the regression model. Therefore, we propose to utilize Ant Colony Optimization (ACO) algorithm for parameter selection of SVR.

In this paper, we build a multi-scale network traffic prediction model based on SVR. First, we use Empirical Mode Decomposition (EMD) [10] to reduce the complexity of data sequence, and decompose the original signals into independent Intrinsic Mode Functions (IMF), so that the network traffic can be divided as multiple multi-scale stable IMF. Then, for each IMF, a SVR based model improved by chaos theory and ACO, called SVR-CACO, is built. Specifically, the parameter selection of SVR is based on ACO algorithm. Also, in order to improve the diversity of solution space and generalization of search space, chaos theory is considered for pheromone update.

The remains of this paper are organized as follows. Section 2 provides some related work. In Section 3, we present our proposed model for network traffic prediction. Empirical experiments are conducted in Section 4. Finally, the paper is concluded in Section 5.

2. Related Work

There exist plenty of works on network traffic prediction. Generally, there are two kinds of models: statistical based method, such as FARIMA model, Markov model, and ARIMA model; and intelligent algorithms based method.

Wen *et al.* [11] proposed a prediction method based on alpha smooth information to deal with the long-range burst prediction problem with FARIMA model. Liu *et al.* [12] utilized wavelet technique for VBR traffic prediction. Xu *et al.* [13] proposed a wavelet forecasting methods by decomposing unstable time series using wavelet decomposition, and then the results are synthesized for prediction the original traffic flow. Qiao *et al.* [14] decomposed the network flow as scaling coefficient and wavelet coefficient, and then employed ARIMA model for prediction. Crouse *et al.* [15] used a Markov model based method to determine the wavelet coefficient. Chen *et al.* [16] first perform preprocessing on the network flow data by wavelet decomposition and reconstruction, and then use FARIMA for prediction. Gao *et al.* [17] proposed to remove the long term dependence based on the experiences using ARMA model and then predict.

Park *et al.* [18] proposed a bilinear regression neural network to improve the accuracy of prediction. Vieira *et al.* [19] proposed a BDF model to model the network traffic based on multi-fractal analysis, and training using adaptive clustering algorithm. Doulamis *et al.* [20] designed an adaptive neural network by dynamically adjusting the weights. Huang *et al.* [21] tried to introduce Genetic Algorithm (GA) for network traffic prediction, and use GA to improve the BP neural network based prediction method. Li *et al.* [22] proposed to perform wavelet decomposition first, and then use ACO to train the parameters of neural network for prediction.

In order to improve the accuracy of prediction, Li [23] utilized multiple models together for prediction, and then return the weighted summary as the final results of prediction. Yao *et al.* [24] combine the network flow prediction and error prediction together using grey model and Markov model *et al.* [25]. We first decomposed the network flow into a multi-scale sequence using wavelet decomposition, and then for each one use a specific prediction model. Lastly, the results are synthesized to predict the original network flow.

3. Network Traffic Prediction Model

Generally, there are three steps in our proposed network traffic prediction model. First, we utilize EMD to decompose the original data for de-noising and stabilizing network traffic data flow into a series of IMFs. Then, for each IMF, construct a SVR based model improved by chaos theory and ACO algorithm named as SVR-CACO for prediction. Last, individual results from each IMF are assembled by SVR to predict for the original network traffic. Figure 1 illustrates the overall structure of our proposed model.

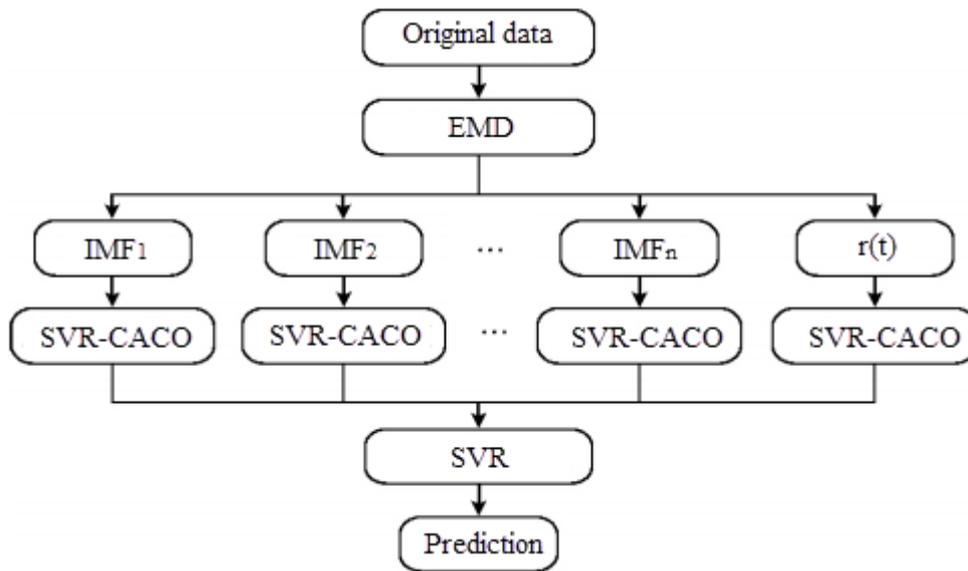


Figure1. Overall Structure of the Proposed Model

3.1. Empirical Mode Decomposition (EMD)

EMD is an established method to deal with nonlinear unstable signals. In this paper, we employ EMD to remove noises and reduce the data complexity. The frequency of signal in EMD is defined upon the locality and instantaneity of the wavelet, so that the instant frequency can be inferred from the changes between time point of signal data, without the existence of fluctuation period.

Original signals are decomposed into several independent IMFs, which satisfy the following rules: (1) the numbers of local extreme points and zero crossing points should be equal or the difference should be not more than one; (2) at each time point, the average of the envelopes of local maximum and local minimum should be zero [26].

Suppose the signal is $X(t)$, and it can be decomposed as a finite number of IMF c_j and a trend term r_n :

$$X(t) = \sum_{j=1}^n c_j(t) + r_n(t). \quad (1)$$

The Standard Deviation (SD) is calculated as:

$$SD(i) = \sum_{i=0}^N \frac{(h_{j,i-1}(t) - h_{j,i}(t))^2}{(h_{j,i-1}(t))^2}, \quad (2)$$

where $h_{j,i-1}(t), h_{j,i}(t)$ are IMFs, j denotes the j -th IMF, i denotes the iteration, and N is the number of iterations. Figure 2 illustrates the process of EMD method.

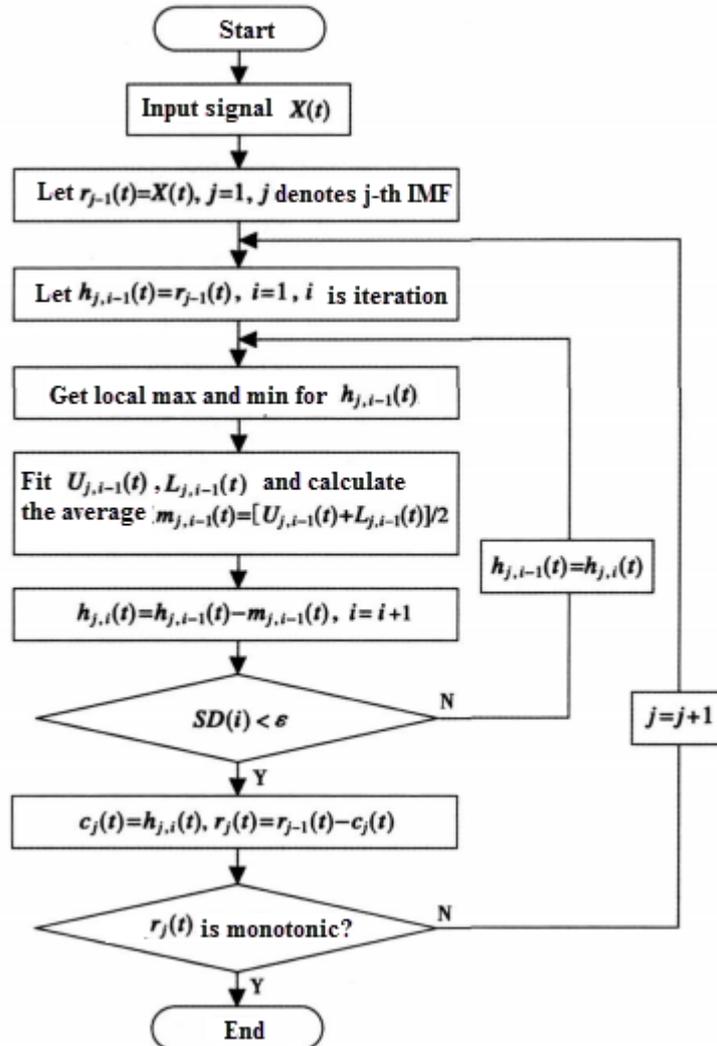


Figure 2. Process Flow of EMD

3.2. Support Vector Regression (SVR)

SVR is a regression method based on SVM. It can be described as follows. Given a set of samples $\{x_i, y_i\}_{i=1}^k$, where x_i is the input value and y_i is the output value. Suppose the linear regression function is $f(x) = w \cdot \Phi(x) + b$, where w is the weight and b is the offset. Parameters w and b can be learned by minimizing the regularized risk function:

$$\min \frac{1}{2} \|w\|^2 + C \frac{1}{k} \sum_{i=1}^k \varepsilon(f(x_i) - y_i) \quad (3)$$

$$s.t. \quad \varepsilon(f(x_i) - y_i) = \begin{cases} |f(x_i) - y_i| - \varepsilon, & \text{if } |w \cdot \Phi(x) + b - y_i| \geq \varepsilon, \\ 0, & \text{otherwise.} \end{cases}$$

where $\varepsilon(\cdot)$ is ε -intensive loss function, and C is the penalty parameter.

Equation (3) can be transformed into the problem of finding the best optimal regression hyperplane:

$$\begin{aligned} \min & \frac{1}{2} \|w\|^2 + C \frac{1}{k} \sum_{i=1}^k (\xi_i + \xi_i^*) \\ \text{s.t.} & \begin{cases} y_i - w \cdot \Phi(x) - b \leq \varepsilon + \xi_i, & \text{if } \xi_i \geq 0, \\ w \cdot \Phi(x) + b - y_i \leq \varepsilon + \xi_i^*, & \text{if } \xi_i^* \geq 0. \end{cases} \end{aligned} \quad (4)$$

where ξ_i, ξ_i^* are non-slack variables.

Introducing Lagrange multipliers a_i and a_i^* , the dual problem is as follows:

$$\begin{aligned} \max & -\frac{1}{2} \sum_{i=1}^k \sum_{j=1}^k (a_i - a_i^*)(a_j - a_j^*) K(x_i, x_j) \\ & - \varepsilon \sum_{i=1}^k (a_i + a_i^*) + \sum_{i=1}^k y_i (a_i - a_i^*) \\ \text{s.t.} & \sum_{i=1}^k (a_i - a_i^*) = 0, C \geq a_i, a_i^* \geq 0 \end{aligned} \quad (5)$$

where $K(x_i, x_j)$ is the kernel function, and $K(x_i, x_j) = e^{-\frac{\|x_i - x_j\|}{\sigma^2}}$, where $\sigma > 0$.

Therefore, the regression we get is:

$$f(x) = \sum_{i=1}^k (a_i - a_i^*) K(x_i, x) + b \quad (6)$$

3.3. SVR Improved by Chaos Theory and ACO (SVR-CACO)

In this section, we introduce our SVR-CACO algorithm, which improves SVR with chaos theory and ACO. Specifically, chaos theory is used to improve the optimal search of ACO, and the modified ACO is used for parameter selection of SVR.

ACO is a swarm intelligence algorithm for optimization problems [27]. The major steps of ACO include next step selection and pheromone update. Let τ_{ij} be the amount of pheromone between cities i and j , and initialized as $\tau_{ij}(0) = B$, where B is a constant. At iteration t , the transfer probability of ant k from i to j is calculated as:

$$p_{ij}^k(t) = \begin{cases} \frac{(\tau_{ij}(t))^\alpha (\eta_{ij}(t))^\beta}{\sum_{u \in allowed_k} (\tau_{iu}(t))^\alpha (\eta_{iu}(t))^\beta}, & \text{if } j \in allowed_k; \\ 0, & \text{otherwise.} \end{cases} \quad (7)$$

where $allowed_k$ denotes the set of qualified cities for next step, α is the heuristic of pheromone indicating the importance of path, β is the heuristic of path. $\eta_{ij}(t) = 1/d_{ij}$ is the expect of ant moving form i to j , where d_{ij} is the distance between cities.

After each ant moves a step or finishes the traversal of all n cities, the pheromone is updated as follows:

$$\tau_{ij}(t+n) = (1-\rho) \cdot \tau_{ij}(t) + \Delta\tau_{ij}(t) \quad (8)$$

where $\rho \in (0,1)$ is the evaporation coefficient of pheromone, and $\Delta\tau_{ij}(t) = \sum_{k=1}^m \Delta\tau_{ij}^k(t)$,

where

$$\Delta\tau_{ij}^k = \begin{cases} \frac{Q}{L_k}, & \text{if ant } k \text{ passes } (i,j); \\ 0, & \text{otherwise.} \end{cases} \quad (9)$$

However, ACO algorithm is easy to fall into local optimum. Considering the characteristics of chaos, such as randomness, ergodicity and sensitivity to the initial conditions, it can traverse all states in a unrepeated way with certain rules. Therefore, we modify ACO with chaos theory to improve the quality of individuals and avoid local optimum.

Introduce Logistic function to construct the chaos system:

$$z_{i+1} = \mu z_i(1 - z_i), i = 0,1,2,\dots, \quad (10)$$

where μ is the control parameter, and $\mu \in (2,4]$. When $\mu = 4$ and $0 \leq z_0 \leq 1$, the system is complete chaos.

The basic idea of utilizing chaos characteristics for optimal search is, to generate a set of chaotic variables, introduce them into optimization process, and extend the search space of chaotic motion as the range of optimal variable, so that the search process is actually conducted by chaotic variables.

The process flow is illustrated in Figure 3, and the steps of SVR-CACO algorithm can be describes as follows.

Step 1: chaos initialization. Randomly generate a λ -dimensional vector $z_1 = z_{11}, z_{12}, \dots, z_{1\lambda}$, where λ is the number of variables in objective optimization function. Then, based on Equation (10), we get n variables z_1, z_2, \dots, z_n , where n is the number of cities.

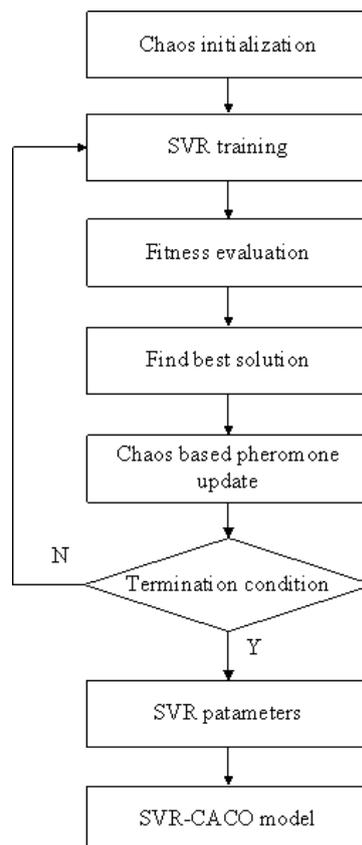


Figure 3. Process Flow of SVR-CACO

Step 2: fitness evaluation based on SVR. Let y_i, \hat{y}_i be the output and predicted value of i -th sample. Calculate the fitness as:

$$fitness = \frac{1}{m} \sum_{i=1}^m \left| \frac{y_i - \hat{y}_i}{y_i} \right|, \quad (11)$$

where m is the number of input samples.

Step 3: find current best solution. Add the fitness value of each ant as a heuristic component into Equation (7) for selecting the next city, and then find the best solution of all ants.

Step 4: update pheromone based on chaos to avoid local optimum by the following equation:

$$\tau_{ij}(t+n) = \rho \cdot \tau_{ij}(t) + \Delta \tau_{ij} + q z_{ij}, \quad (12)$$

where q is the coefficient for chaos, and z_{ij} is chaotic variable.

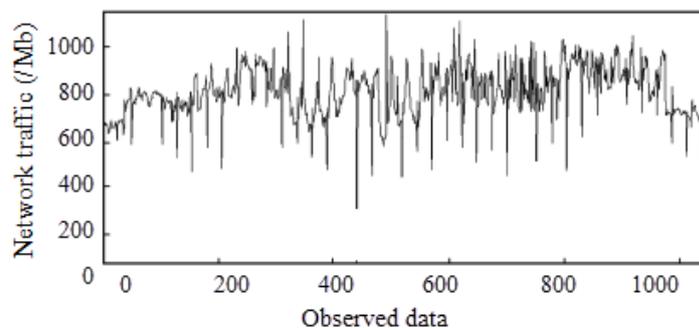
Step 5: if the maximum number of iterations has reached, the algorithm stops, and output the parameters for SVR training; otherwise, return to Step 2.

4. Experiment

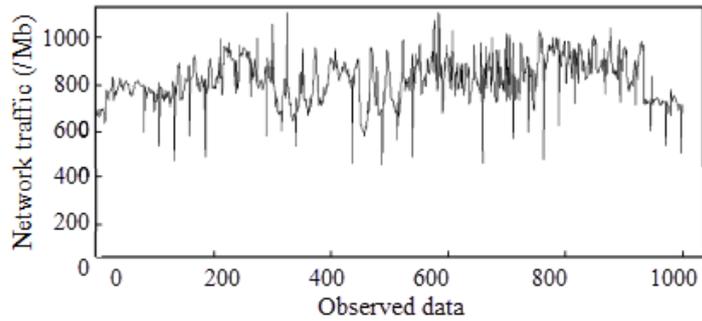
The network traffic data is collect from a broadcasting network center. We collect 6 weeks continuously, 24 hours per day, with the interval of 5 minutes. We have 2,021 observed values per week, and totally 12,126 observed data. We set the time window as 5 minutes, and for each time point, we have 6 values. In real world data measurement, due to the affect of random errors, some bad data should be removed before further processing. We simply perform preprocess based on the deviation of the sample. Let \bar{X} be the average of each data point $x_i (i=1,2,\dots,6)$, and V be the standard deviation, that is:

$$\bar{X} = \frac{1}{6} \sum_{i=1}^6 x_i, V = \sqrt{\frac{1}{6} \sum_{i=1}^6 (x_i - \bar{X})^2}, \quad (13)$$

If x_i satisfies that $|x_i - \bar{X}| > kV$ or $|x_i - \bar{X}| = \max(|x_j - \bar{X}|) (j=1,2,\dots,6)$, then x_i is regarded as bad value and removed. Here k is the coefficient of Grubbs criterion, and we set $k=2.03$. After data preprocessing, we get continuous 1000 data values for prediction. The observed network traffic data is shown in Figure 4.



(a) Original collected data flow



(b) Observed data flow after data preprocessing

Figure 4. Observed Network Traffic Data

We use MATLAB to implement our proposed model. The parameters of SVR optimized by chaos ACO are $C = 180.3824$, $\sigma = 1.0825$, $\varepsilon = 0.0004$, and the relative error is 0.0103. We also compare the parameter selection results of our proposed model (SVR-CACO), original ACO based method (notated as SVR-ACO), and Genetic Algorithm based method (notated as SVR-GA).

Table 1. Comparison of SVR Parameter Selection

Algorithm	C	σ	ε
SVR-CACO	180.3824	1.0825	0.0004
SVR-ACO	145.4732	1.7853	0.0026
SVR-GA	278.5752	2.0661	0.0083

Besides, Figure 5 illustrates the prediction of next 1 hour after training, and Figure 6 is the prediction of next 2 hours.

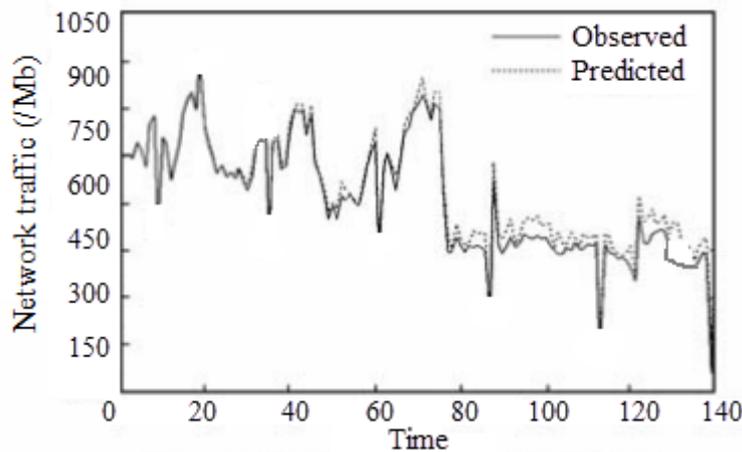


Figure 5. Prediction Results for the Next 1 Hour

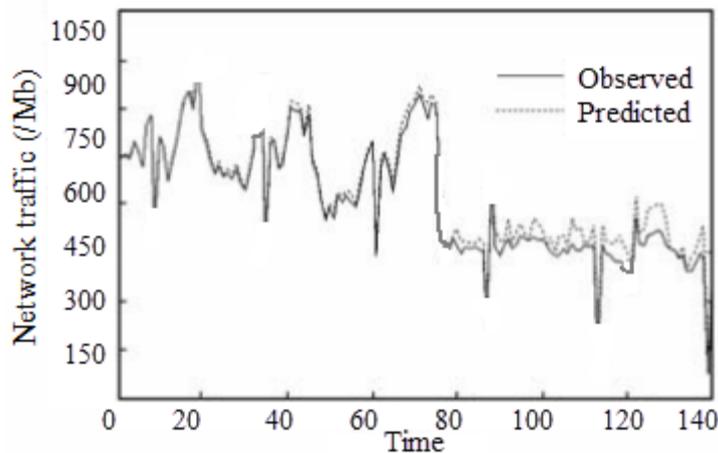


Figure 6. Prediction Results for the Next 2 Hours

5. Conclusion

In this paper, we study on the problem of predicting network traffic. Considering the characteristics of network flow, such as randomness, dynamics and self-adaption, a synthetic method is proposed to predict network traffic. Specifically, we propose to multi-scale method based on EMD, and then apply SVR-CACO for each IMF, and after that, an ensemble SVR is performed for prediction. Besides, our experiments evaluate the efficiency of our model.

This work indicates the feasibility of combining swarm intelligence algorithms with other statistical methods. In future works, we'll try to explore the possibility of combining others for more interesting applications.

Acknowledgements

Our work is supported by the National natural science foundation of China (No. 61103161), the Program for New Century Excellent Talents in University (NCET-12-0579) and the "985" special funds in School of information engineering, Minzu University of China. The authors would like to recognize the others who helped them and thank all our reviewers.

References

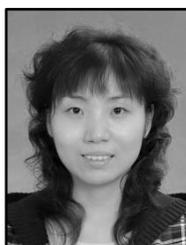
- [1] L. Jinghu, L. Junan and C. Shihua, "Chaotic time series analysis and its applications", Wuhan University Press, (2002).
- [2] Shu, Yantai, *et al.*, "Traffic prediction using FARIMA models", Communications, 1999, ICC'99, IEEE International Conference on, vol. 2, (1999).
- [3] L. B. Smith and M. J. Demetsky, "Short-term traffic flow prediction: neural network approach", Transportation Research Record, vol. 1453, (1994).
- [4] Q.-F. Yao, *et al.*, "Novel network traffic forecasting algorithm based on grey model and Markov chain", Journal of Zhejiang University - Science Edition, vol. 396, (2007).
- [5] Feng and Huifang, *et al.*, "SVM-based models for predicting WLAN traffic", Communications, 2006, ICC'06, IEEE International Conference on, vol. 2, (2006), IEEE.
- [6] J. Ming, C. Wu and D. Hu, "Research on the comparison of time series models for network traffic prediction", Acta Electronica Sinica, vol. 37, (2009), pp. 11, 2353-2359 (in Chinese).
- [7] W. Junsong, *et al.*, "Prediction of internet traffic based on Elman neural network", Control and Decision Conference, CCDC'09, Chinese, (2009).
- [8] M. Hualin, C. Li and L. Zhang, "Network traffic prediction based on grey model and adaptive filter", Computer Engineering, vol. 35, no. 1, (2009), pp. 155-157.
- [9] H. Drucker, *et al.*, "Support vector regression machines", Advances in neural information processing systems, vol. 9, (1997), pp. 155-161.

- [10] G. Rilling, P. Flandrin, and P. Goncalves, "On empirical mode decomposition and its algorithms", IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing NSIP, vol. 3. (2003).
- [11] W. Yong and G. Zhu, "Prediction of self-similar network traffic with heavy tailness", Journal of Huazhong University of Science and Technology (Nature Science), vol. 34, no. 9, (2006).
- [12] H. Liu and G. Mao, "Prediction algorithms for real-time variable-bit-rate video", Communications, Asia-Pacific Conference, (2005).
- [13] K. Xu, J. Xu and X. Ban, "Forecasting of Some Non-Stationary Time Series Based on Wavelet Decomposition", Acta Electronica, vol. 29, no. 4, (2001).
- [14] Y. Qiao, J. Skicewicz, and P. Dinda, "An empirical study of the multiscale predictability of network traffic", High performance Distributed Computing, Proceedings, 13th IEEE International Symposium, (2004).
- [15] S. M. Crouse and R. G. Baraniuk, "Simplified wavelet-domain hidden Markov models using contexts", Acoustics, Speech and Signal Processing, 1998, Proceedings of the 1998 IEEE International Conference on, vol. 4, (1998), IEEE.
- [16] X.-T. Che and J.X. Liu, "Network traffic prediction based on wavelet transformation and FARIIMA", Journal of Communications, vol. 32, no. 4, (2011).
- [17] B. Gao, Q.-Y. ZHANG, Y.-S. Liang, N.-N. Liu, C.-B. and N. -T. Zhang, "Predicting self-similar network traffic based on EMD and ARMA", Journal of Communications, vol. 32, no. 4, (2011).
- [18] D. C. Park, "Structure optimization of BiLinear Recurrent Neural Networks and its application to Ethernet network traffic prediction", Information Sciences, vol. 237, (2013), pp. 18-28.
- [19] Vieira, H. T. Flávio, and G. C. R. Flávio, "An adaptive fuzzy model using orthonormal basis functions based on multifractal characteristics applied to network traffic control", Neurocomputing, vol. 74, no. 11, (2011), pp. 1894-1907.
- [20] A. Doulamis, N. Doulamis, and S. D. Kollias, "An adaptable neural-network model for recursive nonlinear traffic prediction and modeling of MPEG video sources", Neural Networks, vol. 14, no. 1, (2003), pp. 150-166.
- [21] J.-G. Huang, L. Hang, H.-J. Wang and B. Long, "Prediction of Time Sequence Based on GA-BP Neural Net", Journal of University of Electronic Science and Technology of China, vol. 38, no. 5, (2009).
- [22] D.-D. Li, R.-T. Zhang, C.-C. Wang and D.-P. Xiao, "A New Network Traffic Prediction Model Based on Ant Colony Algorithm in Cognitive Networks", Acta Electronica Sinica, vol. 39, no. 10, (2011).
- [23] T.-F. Lin, "A Combined Model for Network Traffic Forecasting Based on Maximum Entropy Principle", Microelectronics and Computer, vol. 23, no. 8, (2006).
- [24] Q.-F. Yao, C.-F. Li, H.-L. Ma and S. Zhang, "Novel network traffic forecasting algorithm based on grey model and Markov chain", Journal of Zhejiang University – Science Edition, vol. 34, no. 4, (2007).
- [25] Y. Wei, J. Wang, C. Wang and K. Zhang, "Network Traffic Prediction Algorithm Based on Wavelet Transform and Combinational Models", Journal of Northeastern University (Natural Science), vol. 32, no. 10, (2011).
- [26] X. Zhang, K. K. Lai, and S.-Y. Wang, "A new approach for crude oil price analysis based on empirical mode decomposition", Energy Economics, vol. 30, no. 3, (2008), pp. 905-918.
- [27] M. Dorigo and M. Birattari, "Ant colony optimization", Encyclopedia of Machine Learning, (2010), pp. 36-39.

Authors



Yonglin Liang. He received his Master of Software Engineering in the college of Software from Huazhong University of Science and Technology. Since 2009 he is a lecturer of Shaoguan University. His current research interests include Artificial Intelligence and Intelligence Human Computer Interaction.



Lirong Qiu. She received her Ph.D. in Computer Sciences (2007) from Chinese Academy of Science. Now she is an associate professor of computer sciences at Information Engineering Department, Minzu University of China. Her current research interests include different aspects of natural language processing, artificial intelligence and distributed systems.