Application of Nonlinear Combination Prediction Model for Network Traffic

Aitao Zhao[1, a], Yingchun Liu[2,b]

[1]Shijiazhuang Vocational Technology Institute , Department of Information Technology
Shijiazhuang,China

[2]Shijiazhuang Vocational Technology Institute, Modern Education Technology Center
Shijiazhuang,China

[a]ait_z@163.com, [b]yingchunliu@139.com

Keywords: Network traffic; Support vector machine (SVM); Prediction model; Combination

Abstract. To improve the network traffic prediction result, the paper put forward a nonlinear combination prediction model of network traffic flow. First, the single models of ARIMA, ARMA, GM (1, 1) were used in the prediction of characteristics of network traffic, Then three predicting results were input to a support vector machine for data fusion to obtain the final forecasting result. The simulation results show that, compared with the other network traffic prediction models, the nonlinear combination forecasting model can well reflect the complex changes in network traffic, thereby improves the prediction accuracy of network traffic.

Introduction

With the rapid development of network technology and the popularization, the network have been tens of thousands of households, the network user excursion, the network service is more and more, network management increasingly heavy work, the network flow control is the basis of network management. Network flow research mainly include network flow modeling prediction and network flow characteristic analysis of two research direction, it is network traffic prediction for large-scale network planning and design has the extremely vital significance, become the current network in the research field of the important topic.

The principle of network traffic combination forecast

Network traffic prediction mathematical model. Network traffic prediction is based on the past. Now we know the network flow history data, establishing a extension from the past to the future network traffic prediction model for network planning, and provide basis for decision-making. Network flow is effected by economy, network users of Internet behavior, weekends and holidays and so on, it has the trend of periodic and nonlinear change rule. Set a length for n of the traffic flow time series x(t), t = 1, 2,... , n, then it can be broken down into the sum of the followings:

$$x(t) = f(t) + g(t) + h(t) + \varepsilon(t) \tag{1}$$

f(t) stands for network flow trend item, g(t) stands for network flow cycle items, h(t) stands for network flow of the nonlinear term, $\varepsilon(t)$ stands for random item.

The choice of a single forecast model. From equation (1), it is known that for the non-linear, periodic and trend of complex network traffic data, using the single forecast model can't the change rule to determine prediction, appropriate mixture of various model combination forecast. However, in the network flow of the combination forecast process, first choose single model, it is also the most key steps, according to the characteristics of the changes of network traffic, choose three ARIMA model of network traffic is a trend to carry on the forecast, the network flow of the ARMA periodic item in prediction, nonlinear term by using grey forecast model, so that the model of the items of the network traffic prediction:

$$\hat{f}(t), \hat{g}(t) \text{ and } \hat{h}(t).$$

The principle of network traffic combination forecast. Due to the network traffic prediction results of the relationship between the items is nonlinear, small sample, and SVM is suitable for small sample, nonlinear prediction problem, so using SVM can effectively decide the relationship between items, so as to improve the network traffic prediction accuracy. Specific combination principle is:

$$\hat{x}(t) = f(\hat{f}(t), \hat{g}(t), \hat{h}(t)) \tag{2}$$

f(x) stands for SVM prediction model.

Network flow combination forecast model

Spheroidizing treatment of network flow data. Because of the network flow by economy, network users of Internet behavior, weekends and holidays and so on the many kinds of factors influence, change amplitude is quite large, but also has a strong trend, and at the same time, the network flow collection all is according to the certain time interval for acquisition, is one kind has the strong trend of time sequence, so need to carry on the normalization processing. Specific is:

$$x_i' = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{3}$$

The result of prediction in reverse normalization formula is:

$$x_i = x_i' \times (x_{max} - x_{min}) + x_{min} \tag{4}$$

x_i is the original network traffic data. x_i' is the data after the normalization of network flow. x_{max} and x_{min} are represented the network flow's upper bound and lower bound.

The trend of the network traffic prediction model. For some non-stationary, with periodic time series $\{x_t\}$, it can use the difference method and the incremental sequence $\nabla x_t = x_t - x_{t-1}$, abate the relationship between network traffic sequence, make its become more smoothly, through d order difference, incremental sequence $\{\nabla^d x_t\}$ become stationary series. Then according to the relationship of difference and retrusion operator, we can get the ARIMA model:

$$\varphi(B)(1-B)^d x_t = \theta(B)\alpha_t \tag{5}$$

d is difference operation order. n and m is autoregressive and moving average order number.

The specific modeling and forecasting process of ARIMA network flow time series prediction is showed in figure 1.
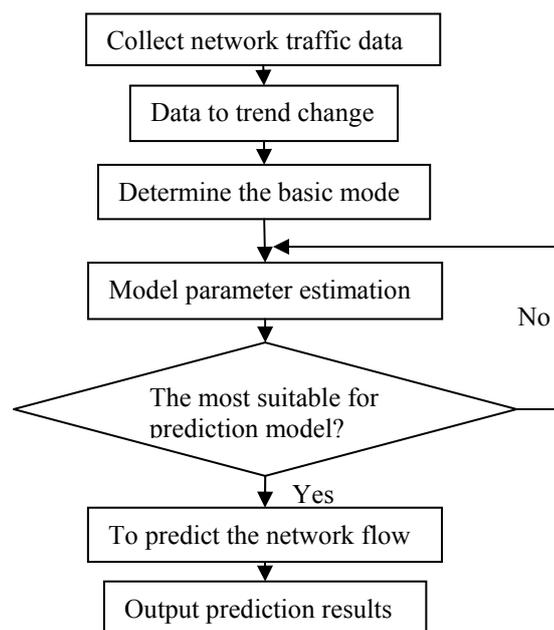

Figure 1. Network flow trend term prediction process

The cycle of network traffic prediction model. ARMA model is by autoregressive model (AR) development, and is very suitable for has periodic time series prediction, so using the network traffic

prediction of periodic law. Set to collect network flow time series $\{x_t\}$ is smooth and zero mean sequence, then establish fitting difference equation model for $\{x_t\}$:

$$x_t - \varphi_1 x_{t-1} - \varphi_2 x_{t-2} - \cdots - \varphi_n x_{t-n} = \varepsilon_t - \theta_1 \varepsilon_t - \cdots \theta_m \varepsilon_{t-m} \tag{6}$$

Simple denoted:

$$x_t - \sum_{i=1}^{n} \varphi_i x_{t-i} = \alpha_t - \sum_{j=1}^{n} \theta_j \alpha_{t-j} \tag{7}$$

When m = 0 and n = 0, the ARMA model into the AR model and the MA model points, so it's quite flexible, in equation (6) introducing retrusion operator B, then:

$$(1 - \varphi_1 B - \varphi_2 B^2 - \cdots \varphi_n B^n) x_t = (1 - \theta_1 B - \theta_2 B^2 - \cdots \theta_m B^m) \varepsilon_t \tag{8}$$

The simulation research

Data sources. Experimental data from the network flow library: HTTP://newsfeed.ntcu.net/~ news|2006|, collect the Lord node routers from Incoming articles on February 1, 2011 to March 20, network access flow per hour, the acquisition of 800 data, form a network traffic time series $\{x (t), t = 1, 2,... ,800\}$, before 750 data as training set to establish network traffic prediction model, after 50 data as a test set forecast test. The network flow original data normalization processing.

model performance evaluation index. In order to assess the model performance, the root mean square error (root mean squared predict error, RMSE) and mean absolute relative error (mean absolute percent error, MAPE) as a model of evaluation index.

ARIMA model of the network traffic prediction. Using ARIMA to network flow forecast, get the optimal prediction mode is: ARIMA (1,1,1). ARIMA model of the network traffic prediction accuracy is low, but can be very good prediction network flow trend change rule, therefore USES a single ARIMA model is not suitable for complex, nonlinear network traffic prediction.

ARMA model of the network traffic prediction. Used to forecast the ARMA network traffic, and get the optimal prediction mode is: ARIMA (3, 2). ARMA model of the network traffic prediction the same low accuracy, can only can describe network traffic periodic change rule, using single ARIA model is not suitable for complex network traffic prediction.

GM (1, 1) model for the network flow prediction. Used ARMA model for forecasting network traffic, and get the optimal prediction mode is: ARIMA (3, 2). GM (1, 1) can be very good prediction of nonlinear network flow change rule, but the prediction accuracy and the actual network flow value far, a single use GM(1, 1) for complex network traffic forecast results are not reliable.

Combination ARIMA, ARMA and GM (1, 1). The ARIMA, ARMA and GM (1, 1 prediction) input to the SVM as the training sample input, the last 50 network flow observations as expected output, using the grid search method of SVM parameter optimization, and USES the leave a method for cross validation to select the most optimal prediction model, finally, the network traffic forecast test set.

Contrast with other prediction models. In order to make the prediction model prediction results more convincing, using single ARIMA model and ARMA, GM (1, 1) and the combination model ARIMA - ARMA, ARMA - GM (1, 1), ARMIA - GM (1, 1) as a reference model, the result of the test by the RMSE and MAPE comparison, and the specific contrast results as is shown in Table 1.

Table 1. Several kinds of prediction model forecast performance contrast

Model	RMSE	MAPE(%)
ARIMA	4. 24	9. 52
ARMA	3. 44	9. 67
GM(1，1)	4. 21	8. 86
ARIMA－ARM	2. 44	3. 16
ARMA－GM(1，1)	3. 57	6. 35
ARMIA－GM(1，1)	2. 45	3.45
Non-linear combination forecast model in this article	1.09	1. 96

From table 1 contrast result, the single forecasting model (ARIMA, AR - MA, GM (1, 1) prediction accuracy is lower than simple combination model that combined model can more reflect the change rule of network traffic, and the proposed nonlinear combination forecasting model is far higher than other contrast combination the prediction precision, the network traffic prediction error minimum, compared to the experimental results show that the proposed nonlinear network flow prediction model can comprehensive, accurate describe network traffic trend, periodic and nonlinear change rule, prediction results more credible and reliable, prediction results can help network manager better understand the current network running status.

Conclusions

In view of the current single network traffic, and the simple combination forecast model to depict existing network traffic complex change characteristics, combining with the ARIMA, ARMA and GM (1, 1) of different advantages, first had to use them to all kinds of network flow change law to carry on the forecast, then the small sample, nonlinear SVM to three forecast results are not line will sex fusion, for the network traffic prediction results. The simulation results show that the proposed nonlinear network flow combination forecast model to improve the network traffic prediction accuracy, in the field of network management has broad prospects for application.

References

[1] Park J H. Synchronization of Genesio Chaotic System via Back- stepping Approach[J]. Chaos Solitons Fractals, 2006, 27(5): 1369- 1375.

[2] Yan Junjun, Yang Yi-Sung, Chiang Tsung-Ying, et al. Robust Synchronization of Unified Chaotic Systems via Sliding Mode Control[J]. Chaos, Solitons and Fractals, 2007, 34(3): 947-954.

[3] Julisch, K. Mining Alarm Clusters to Improve Alarm Handling Efficiency. in: Proceedings 17th Annual Computer Security Applications Conference. New Orleans, LA, USA: IEEE Comput. Soc, 2001. 12.

[4] Goodall, J. R., Lutters, W. G., Komlodi, A. The Work of Intrusion Detection:Rethinking the Role of Security Analyst. in: Proceedings of the Tenth Americas Conference on Information Systems. New York: 2004.

[5] Ammann, P., Wijesekera, D., Kaushik, S. Scalable, Graph-based Network Vulnerability Analysis. in: Proceedings of 9th ACM Conference on Computer and Communications Security (CCS 2002): 2002.

[6]Yanmei Li, Jingmin Wang. The Load Forecasting Model Based on Bayes-GRNN. Journal of Sof tware. Vol.7, No.6, 2012,6 :1273～1280.