Nonlinear-Periodical Network Traffic Behavioral Forecast on Seasonal Neural Network Model[1]

Cheng Guang, Gong Jian, Ding Wei
gcheng@njnet.edu.cn
Department of Computer Science & Engineering, Southeast University (Nanjing City 210096)

Abstract: How to predict Internet behavior is a standing challenge to the study of network behaviorism. Traditionally, the ARIMA model, always used to network traffic prediction, has difficulties in decision its parameter values and therefore, is hard to dealing with condition of Non-linear time series. This paper represents a seasonal neural network prediction model on monitoring network traffic based on the neural network model of time series and the periodic trend of traffic behavior. What's more, a series of data processes are taken to improve the accuracy of prediction. The result from the application of the model on CERNET traffic prediction shows the model's reasonableness, and that it's more accurate than the ARIMA model and common time series of neural network.

Keywords: Traffic Prediction, Seasonal Neural Network, Non-linearity

1. Introduction

Network traffic prediction is being wildly used in many fields, such as dynamic bandwidth allocation schemes for VBR video, network planning for wide area networks, predictive congestion control and so on. In 1994, Nancy and George [1] used time series analysis on detailed forecasting NSFNET backbone traffic. And ARIMA model made the forecasts year in advance at traffic level. In 1995, Sabyasachi and Amarnath [2] modeled data traffic sequences as ARMA processes. In 2001, SHU Yan-Tai[3] provided ways to predict network traffic with FARIMA model.

The parameter set of a standard model, such like the ARIMA, used to be derived from the auto-correlation and frequency spectrum of the time series. When using the ARIMA model to represent time series nonlinear processes problems arise. A number of recently experiences of studies on traffic measurements [4] have convincingly demonstrated that actual network traffic behavior exhibits the Noah Effect (i.e., having high variability or infinite variance), and aggregated network traffic shows the Joseph Effect (i.e., being self-similar or long-range dependent).

The usage of artificial neural networks (ANN) on traffic analysis relies mostly on the measured data. As multi-layer feed forward networks with at least one hidden layer and a sufficient number of hidden units are capable of estimate any measure function, an ANN is powerful enough to represent time series of network traffic. Even In the case of noisy and/or missing data, the generalization makes ANN to work properly. Another advantage of ANN over linear models like the ARIMA approach is the ability to represent nonlinear network traffic [5]. Based on ARIMA theory and ANN theory, a dynamic seasonal TS NN prediction model is proposed in this paper. In the model, the periodic changing characters of traffic for a large-scale network in long time-series are revealed sufficiently. To improve the prediction precision the influence of bizarre data is eliminated, and smoothing process on monitoring data is proceed so as to avoid the losing of series periodicity. In general models, when the prediction step increases, more prediction errors will be involved. However, in this model, a period of T-step prediction can be obtained by a single calculation. The proposed model is discussed in detail in the paper, and is used for traffic prediction on CERNET Eastern China (north) regional network, a large-scale network with more than 500,000 users.

2. Seasonal Neural Network Model

[1] This research was partially supported by 863 program of China under grant No. 2001AA112060 and NNSFC under grant No. 90104031.

2.1 Traditional time-series neural network model

The normal TS NN prediction models can be distinguished as two types. One is a NN on single-variable TS, and can be shown as following. If a TS $X=\{X_1, X_2, \cdots, X_n\}$is resumed, the future value of X has a functional relation with its forward m values. The function is denoted as

$$X_{n+k}= F(X_n, X_{n-1}, \cdots, X_{n-m+1}) \quad (1)$$

Then the functional relation F(•) is fit by adopting the NN, and the predicting value can be obtained.

The other is a NN model on multi-variable TS shown as $(X_{1,1}, X_{2,1}, \cdots, X_{p,1})(X_{1,2}, X_{2,2}, \cdots, X_{p,2})$, including p variables. Same as single variable, the relation F(•)between future value and their previous m values is denoted as the following

$$(X_{1,n+k}, X_{2,n+k},\cdots,X_{p,n+k})= F[(X_{1,n-m+1},X_{2,n-m+1},\cdots,X_{p,n-m+1}),\cdots,(X_{1,n},X_{2,n},\cdots,X_{p,n})] \quad (2)$$

The function is fit by NN and the predicting procedure can be executed.

2.2 The features of network traffic

The TS of network traffic X(t) is made up with four parts including trend part, period part, mutation part and random part, calculated as $X(t)=A(t)+P(t)+B(t)+R(t)$. Some other science areas also use the assumptions, for example, the time-series of hydroaraphy is consisted of the four parts [8]. Owing to the influence of network users and the applications, the trend item A(t) indicates seasonal or many-year changing trend of the traffic caused by the technical innovation, the network size increase, and the application changes as well. The period item P(t) reflects the periodic changes of the traffic. The mutation item B(t) is the burst part of the traffic caused by e.g. network fault or misuse. The trend item, the period item, and the mutation item reflect the ascertainable parts in the traffic TS. The rest part in the traffic is random item R(t).

From the long TS of a network traffic, it may show daily, weekly, and yearly periodic changing characters, so P(t), B(t) and R(t) will be processed separately in this paper. In the raw data treatment procedure, after B(t) is erased, data smoothing process is taken to eliminate R(t), then A(t) and P(t) are left.

Owing to the limitation of the training sets, the TS X'(t) of NN output will produce random item R'(t) again, so X'(t) is made up with three parts including trend part, period part, and random part, calculated as $X'(t)=A'(t)+P'(t)+ R'(t)$. After taken off the random item R'(t) from X'(t) with the data smoothing process again, the predicted time series X"(t) can be calculated as $X''(t)=A'(t)+P'(t)$.

2.3 The model structure

According to the discussion above, the seasonal neural network model (SNN) can be determined by the following procedure: The seasonal TS is assumed as X: $X_1, X_2, \cdots, X_i, \cdots$, and S is its period. That is to say, X can be denoted by $X=\{(X_{1,1}, X_{1,2}, \cdots, X_{1,s}), (X_{2,1}, X_{2,2}, \cdots, X_{2,s}), \cdots, (X_{i,1}, X_{i,2}, \cdots, X_{i,s}), \cdots\}$. In the model, it is assumed that the future period has a functional relation with the historical or past d+1 periods value. The function can be defined as the following equation.

$$(X_{t+d+1,1},X_{t+d+1,2}, \cdots,X_{t+d+1,s}) =G(X_{t,1}, X_{t,2}, \cdots, X_{t,s}; X_{t+1,1}, \cdots, X_{t+1,s}; \cdots; X_{t+d,1}, X_{t+d,2}, \cdots, X_{t+d,s}) \quad (3)$$

the periodic function G(•) is fit by using NN, and then is used for the prediction of future period values. The structure of the SNN is illustrated as Fig.1.

The model consists of three layer nerves, called input layer (IL), hidden layer (HL), and output layer (OL) from top to bottom. The nerve cells of IL map the discrete points from t^{th} to $(t+d)^{th}$ period of periodic TS. The nerve cells of OL map the discrete points of $(t+d+1)^{th}$ period. The main idea of model is preferred as the following.

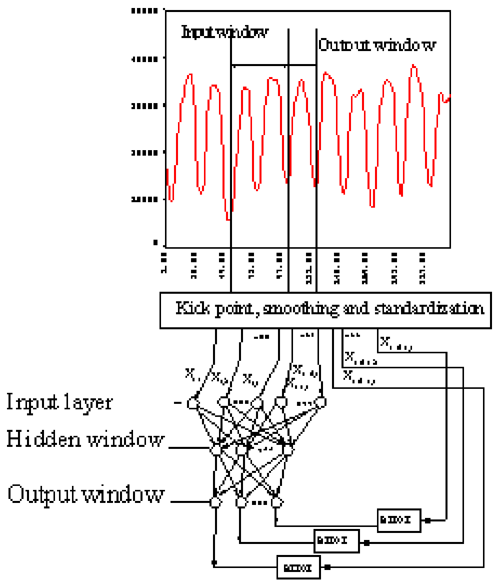


Fig.1 Structure of the SNN Model

For a seasonal TS X with period S, it is assumed that X contains m series X_1, X_2, …, X_m with period S. In this way, we can obtain that the model has (d+1)×s nerve cells in IL and s nerve cells in OL. So the selected n learning samples can be shown as $P=(P_1=(X_1, X_2, \cdots, X_s); P_2=(X_2, X_3, \cdots, X_{s+1}); \cdots; P_n=(X_n, X_{n+1}, \cdots, X_{n+s-1}))$, $T=(T_1=X_{s+1}; T_2=X_{s+2}; \cdots, T_n=X_{s+n})$ .The learning purpose is to adjust the weight values according to the errors between the actual output (A_1, A_2, …, A_n) of NN relevant with P_1, P_2, …, P_n and the factual values T_1, T_2, …, T_n for n training samples. The weight value is adjusted by the back propagation of error signal along original connection path until the expected precision is obtained, i.e. output value A_i and expected T_i (i=1, 2, …, n) differs little. When the square error of OL reaches the satisfied precision, the training process will be terminated.

2.4 Data Treatment

2.4.1 Eliminating Bizarre Data

In order to avoid the influence of bizarre data produced by contrived factors in gathering data, and avoid their influence in predicting process, bizarre data should be examined and eliminated before setting up the model. So in the paper two data choke filters whose output is the smoothing estimation of the input function are introduced in the advanced model [7].

It's supposed that $(\overline{X_t})^2$ is the square value of the data smoothed, with $\overline{X_t^2}$ smoothed value of data square, and sample variance $S^2(t)$ is defined as $S^2(t)=\overline{X_t^2}-\overline{X_t}^2$,with standard deviation S(t) as the square root of sample variance $S^2(t)$, then the next data X_{t+1} can be examined. If the condition $\overline{X_t}-kS(t)<X_{t+1}<\overline{X_t}+kS(t)$ is satisfied, X_{t+1} can be accepted as the useful information. Its constant parameter k varies from 3 to 9 based on the traffic sequence and its original value is adopted 6 commonly[7]. Otherwise X_{t+1} is considered as bizarre data, and X_{t+1} is substituted by $\hat{X}_{t+1}$, which is defined by $\hat{X}_{t+1}=2X_t-X_{t-1}$.

2.4.2 Smoothing Treatment Process

A long-term TS of observed data contain a certain white noise character, so a smoothing process is required in TS to erase the influence of white noise to some degree. In the paper, the proposed model uses 'Fourier analysis method' [7] to obtain the smoothed function $\hat{X}_d$ of X_d so as to decrease the error. Before using the method the smoothing coefficient should be ensured firstly. According to Fourier function theory, the smoothing value of the time series is defined as X_d,

$$X_d=\overline{X}_d+\sum_{j=1}^{h}[A_{j,d}\cos\left(\frac{2\pi jt}{N}\right)+B_{j,d}\sin\left(\frac{2\pi jt}{N}\right) \tag{4}$$

Where $\overline{X}_d$ is the average value of d^{th} day traffic and it is defined as $\overline{X}_d=\frac{1}{N}\sum_{t=1}^{N}X_{t,d}$. If N is even number then h=N/2, otherwise h=(N-1)/2. $A_{j,d}$ and $B_{j,d}$ are Fourier coefficient or smoothing coefficient and they are defined by

$$A_{j,d}=\frac{2}{N}\sum_{t=1}^{N}X_{t,d}\cos\left(\frac{2\pi jt}{N}\right);$$

$$B_{j,d}=\frac{2}{N}\sum_{t=1}^{N}X_{t,d}\sin\left(\frac{2\pi jt}{N}\right)\quad (j=1, 2,\ldots, h) \tag{5}$$

The variance $\mathrm{var}(X_d)=\frac{1}{N}\sum_{i=1}^{N}(X_{i,d}-\overline{X}_d)^2$ reflects the moving range of smoothing value and $\mathrm{var}(j,d)=\frac{1}{2}\left(A_{j,d}^2+B_{j,d}^2\right)$ reflects the influence degree that the j^{th} smoothing coefficients influence the total variance. The model uses the F hypothesis verification method to search the value of j:

$$F_{j,d}=\frac{\mathrm{var}(j,d)}{\mathrm{var}(X_d)/h}=\frac{\mathrm{var}(j,d)}{\mathrm{var}(X_d)}\times h \tag{6}$$

Given trust level α , if $F_{j,d}>F_\alpha(1,12)$, j is considered, else i is rejected. α is chosen as 0.05 commonly.

It is provided that $M = K \times N$, where M, K is separately sampling total number and day number of the total traffic, and N is sampling number measured in one day. According to the equations (4, 5, 6) the TC of $A_{j,d}$ and $B_{j,d}$ is O(N), and the TC of X_d is $O(N^2)$, so the total TC of the algorithm is $O(K \times N^2)$, that is $O(M \times N)$. There is no specific restriction on this method. If the length N of samples is just the multiple times of the period within one day, a precise result can be obtained, otherwise the result will be approximate. The experiment shows that the model is convergent within traffic data measured from the real environment.

3. A Case Study

3.1 Background

The raw traffic data used in the paper comes from CERNET Eastern China (north) regional network. The actual sample monitoring data are shown in Fig.2. By the Figure, the traffic TS has obviously a daily period, because there are more network users in daytime than in nighttime. Due to the Spring Festival on Jan 24th, the traffic in that day is the least. And because after Feb 7th many students came back to campuses, the traffic increased rapidly from Feb 7th to 8th. It is shown that the traffic in the CERNET is dominated by students' behavior.

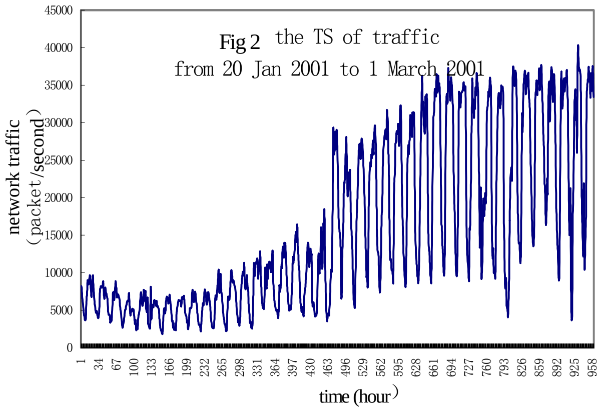
3.2 SNN model structure

The number of input units, equivalent to the size of the input window, determines the number of periods the NN "looks into the past" when predicting the future. The number of output units, which is equivalent to the size of the output window, determines the number of periods the NN that predicts the future [6]. Whereas it has been shown that one HL is sufficient to approximate continuous function. There are nearly 30 days data are used to train the model and the following 10 data are used to calibrate the model. The training network is selected with 48 nerve cells in IL and they map 48 hours network traffic of two days, that is to say, the input of network including two periods with 24 hours traffic. Accordingly, 24 nerve cells in OL map 24 hours network traffic of one day. The network traffic level series after eliciting oddity data and smoothing process are expressed as the following:

$$traffic=\{t_1=(t_{1,1}, t_{1,2}, \cdots, t_{1,24}); t_2=(t_{2,1}, t_{2,2}, \cdots, t_{2,24}); \cdots; t_{40}=(t_{40,1}, t_{40,2}, \cdots, t_{40,24})\} \quad (7)$$

Then the standardized data is taken for training. Learning sample is $P=\{(t_1, t_2); (t_2, t_3); \cdots; (t_{29}, t_{30})\}$, where $T=\{t_3; t_4; \cdots; t_{32}\}$ is corresponding output. To the HL, 64 nerve cells are selected according to many times experiments.

We used the following terms to measure the model performance metrics s_m and forecasting performance metrics s_f of our system: For a TS $X_1, \ldots, X_n$



$$s_m = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \hat{X}_i)^2}{n}}, \; s_f = \sqrt{\frac{\sum_{i=n+1}^{n+1+r}(X_i - \hat{X}_i)^2}{r}} \quad (8)$$

Where $\hat{X}_i$ is the estimation of the ANN for period i and r is forecasting period numbers. The model performance metrics s_m estimates the capability of the NN to mimic the known data set, the forecasting performance metrics s_f judges the forecast capability of the network for a forecast period of length r. In the model r=24 is adopted.

3.3 Traffic Prediction

The model can predict traffic behavior after it has learned. Owing to the limitation of learning sample, the prediction value will bring noise, so the result of traffic prediction must also be smoothed to cancel the noise. It has proved that the smoothed predictive result is better than non-smoothed predictive result. During the prediction of the model, firstly the traffic of 30th day and 31st day, T_{30} and T_{31}, predict the traffic YCT_{32} of 32th day. Secondly T_{31} and T_{32} predict the traffic YCT_{33} in 33th day. The rest may be deduced by analogy until the traffic YCT_{40} is predicted. The final predictive result $SYCT_{32} \ldots SYCT_{40}$ can obtain by smoothed the prediction result of the NN. According to the model built, the predicted network traffic from 20 Feb to 1 Mar is shown as Fig.3. From the prediction analysis figure, we can conclude that the prediction trend and value are all

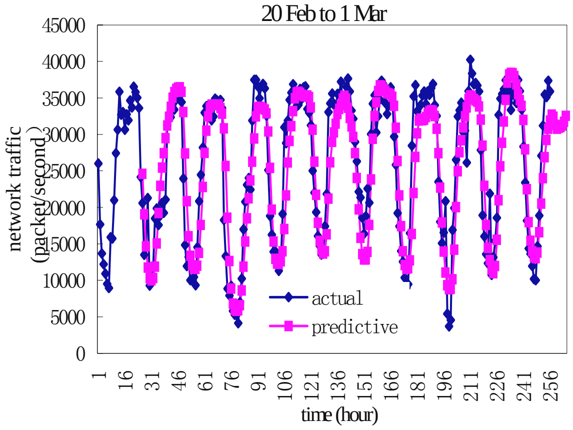


Fig 3 the TS of the predictive value and actual data from 20 Feb to 1 Mar

conformable. The model operated well with much higher prediction precision.

3.4 Comparison with Other Model

Time series	forecasting performance metrics s_f		
	SNN	NN	SARIMA
Feb 28th	944.1	1772.1	1224.2

Table1: forecasting errors for three models on Feb 28th

We compared our results on Feb 28th with the results of the ARIMA model and the NN model. As an opponent for the ARIMA model and the NN model, we selected those networks that delivered the smallest forecasting performance metrics s_f for the respective TS data. s_f for the SNN model, the NN, and the Seasonal ARIMA model (SARIMA) are compared and shown in Table1. The SNN model outperformed the SARIMA models, whereas the NN model was inferior to the SNN model and the SARIMA model. This behavior can be explained as following: the network traffic behavior is obviously daily periodicity and non-linearity, and the NN model doesn't consider the periodicity of traffic, and the SARIMA model can only process the linear random variable.

4. Conclusions

On a large-scale network like CERNET, the traffics show periodic features because of factors. However it is difficult to ascertain the relationship among these factors. So in the paper, based on the TS and NN analysis method, a new seasonal network behavior prediction model is represented and described in details. In the model, it is critical to take period of TS as an input and output unit of the network. The model will avoid lose of periodic information of TS. And more, when the model is operated for only one step, the predicting results of T steps is obtainable. Finally, according to experiment on CERNET, the model works well in practice.

References

[1] N. Groschwitz, G. Polyzos, A Time Series Model of Long-term Traffic on the NSFnet Backbone, In Proceedings of the IEEE International Conference on Communications(ICC'94), May 1994.

[2] S. Basu and A. Mukherjee. Time series models for Internet traffic, INFOCOM '96. Fifteenth Annual Joint Conference of the IEEE Computer Societies. Networking the Next Generation, Proceedings IEEE , Volume: 2 , 1996, pp. 611—620.

[3] Shu yan-Tai, Wang Lei, Zhang Lian Fang, etc. Internet Traffic Modeling and Prediction Using FARIMA Models, Chinese J. Computer, Vol. 24, No.1 Jan. 2001, pp. 46--54. (in Chinese)

[4] V. Paxson, S. Flod, Wide-area traffic: The failure of possion modeling, In Proceedings of the ACM Sigcomm'94, London, UK, 1994, pp. 257—268.

[5] Rich Wolski, Forecasting Network Performance to Support Dynamic Scheduling Using the Network Weather Service.

[6] Thomas Kolarik, Gottfried Rudorfer. Time Series Forecasting Using Neural Networks. ACM, time series & neural networks, Sept. 1994. pp. 86--94

[7] Yang Wei-tsing, Gu Lan. Time-series analysis and dynamic data modeling, BeiJing University of Science and Technology Press, 1988.12, BeiJing. (in Chinese)

[8] Jing Guang-yan, Random analysis of hydroaraphy and hydro-resource, China Science and Technology Press, 1992.5, BeiJing. (in Chinese)