Research on the Network Traffic Time Series Modeling and Forecasting Based on Wavelet Decomposition

Zhijie WANG, Weiping LI

Abstract

A network traffic time series analysis and forecasting method based on wavelet decomposition is proposed in the paper. Firstly, non-stationary network traffic time series are decomposed into multiple stationary components, then various stationary components are modeled using the autoregressive moving average method and last models of all components are composed to get the prediction model for original non-stationary network traffic time series. In the simulation experiment, we used time series data from the network traffic library to establish a prediction model and carried out independent test and inspection. Simulation test results show that the forecasting method proposed in this paper has improved forecasting accuracy. It is an effective and robust network traffic forecasting method.

Keywords: Network Traffic, Wavelet Slicing, Time Series, Forecasting

1. Introduction

Network traffic is generally sampled at a certain time interval. It is a kind of typical time series data. Therefore, time series analysis method has been widely used in network traffic forecasting applications for disclosing inherent statistical characteristics and change law of network traffic [1]. In the actual application process, it is generally assumed in time series analysis that network traffic change is a kind of stationary time series data without much non-stationary network traffic time analysis [2]. However, the network system is very complex and most network traffic has non-stationary characteristics. Therefore, it is of great significance non-stationary time series analysis for discussing non-stationary time series analysis on network traffic analysis and forecasting [3].

In non-stationary network traffic time series forecasting applications, they are often converted into stationary time series, then the stationary time series forecasting method is used for analysis and forecasting, such as: autoregressive moving average model (ARMA), controlled autoregressive model (CAR), autoregressive integrating moving average model (ARIMA) [4-6]. These traditional time series analysis methods belong to linear model, suitable for stationary time series forecasting for seasonal and cyclical features. While for some non-stationary and nonlinear network flow time series, as they cannot fully reflect complex variation characteristics of time series, the accuracy of forecasting is relatively low [7].

Wavelet analysis was proposed by Meyer and Mallat. It can decompose time series data into relatively simple time series [8]. In order to improve accuracy of network traffic forecasting, a network traffic time series analysis and forecasting method based on wavelet decomposition is proposed in this paper. Firstly, the non-stationary network traffic time series are decomposed into multiple stationary components through wavelet, and then the autoregressive moving average method is used to model various stationary components and at last models of all components are combined to get a prediction model of original non-stationary network traffic time series.

2. Network Traffic Modeling Method

2.1 Time Series Wavelet Analytical Solution

By using the multi-resolution function of wavelet analysis for decomposing, time series is decomposed into two parts: low-frequency and high frequency coefficients. Through restructuring operation, low-frequency components of original time series can be obtained from low-frequency coefficients and high-frequency components can be obtained from high frequency coefficient [9].
Time series decomposition can be realized by using Mallat algorithm specifically. See the decomposition relationship as follows:

\[
\begin{align*}
    d_{j+1} &= h_0 * a_j \\
    a_{j+1} &= h_1 * d_j
\end{align*}
\]

Where, \(h_0\) is low pass decomposition filter; \(h_1\) is high pass resolution filter; * is convolution operator; \(a_j\) is low-frequency coefficient and \(d_j\) is high frequency coefficient.

It indicate from formula (1) that, when \(j=0\), \(a_0\) stands for original time series. After many times of decomposition operations of the original time series for \(h_0\) and \(h_1\), low-frequency and high frequency coefficients of the original time series can be obtained.

### 2.2 Reconstruction of Time Series after Decomposition

After wavelet decomposition for the time series, low-frequency and low-frequency coefficients can be obtained, then they are used for restructuring operation to get low-frequency and high-frequency components of the time series. Wavelet reorganization relationship is as follows:

\[
\begin{align*}
    A_j &= g_0 * a_j \\
    D_j &= g_1 * d_j
\end{align*}
\]

Where, \(g_0\) is low-pass reorganization filter; \(g_1\) is high-pass reorganization filter; \(A_j\) is low-frequency component and \(D_j\) is high-frequency component.

Thus the relationship between original time series and low-frequency component \(A_j\) & high-frequency components \(D_j\) is as the following:

\[
S = A_j + D_j
\]

### 2.3 Modeling Steps of Time Series Forecasting Models

1. Suppose data of a time series S is received.
2. Identify its non-stationarity: if it is non-stationary time, use wavelet analysis to decompose, and then get low-frequency and high frequency coefficients of the time series.
3. Use restructuring operation algorithm to reconstruct low-frequency and high frequency coefficients of time series after decomposition to get low-frequency component \(A_1\) and high-frequency component \(D_1\).
4. Re-carry out stationarity test for \(A_1\) and \(D_1\). If a component meets features of stationary time series, it may not be decomposed, or the time series component needs to be further decomposed and reconstructed until it is converted into a stationary time series.
5. If the autocorrelation function of time series components has trailing features and the partial correlation function is truncated in Step p of time delay, use the autoregressive (AR) model to carry out modeling forecasting for the time series components.
6. If the partial correlation function of time series components has trailing features and the autocorrelation function is truncated in Step p of time delay, use the moving average (MA) model to carry out modeling forecasting for the time series components.
7. If the time series is neither suitable for AR model nor for MA model, then adopt ARMA (m, n) model to carry out modeling forecasting. The best exponent number of the model is determined according to BIC (Bayes Information Criterion) or AIC (Akaike Information Criterion) and model parameters are estimated according to the least square method.

### 3. Simulation Experiment

#### 3.1 Data Source

The experimental data is sourced from network traffic library: http://newsfeed.ntcu.net/~news/
where network per hour visitor traffic of the main node router, Incoming Articles from July 1, 2011 until November 10, 2011 and 500 digits were obtained to form a network traffic time series \( \{ x(t), t = 1, 2, \ldots, 500 \} \), of which the first 450 digits are used as training samples for modeling and are fitted as historical data; the latter 50 digits are used as test samples for forecasting and testing of model performance. See Fig.1 for specific data.

### 3.2 Data Analysis

Firstly, autocorrelation function and partial correlation function of the network traffic time series are analyzed to determine stationarity of the primitive network traffic time series, using the analytical software DPS9.0 to get results as shown in Fig.2 and Fig.3. It can be determined from Fig.2 and Fig.3 that autocorrelation function and partial correlation function of primitive network traffic time series have no significant attenuated trend but with oscillatory wave. Therefore, the primitive network traffic time series do not meet features of stationary time series. Using ARMA model for direct modeling, we may get un-reliable results.

![Fig. 1 network traffic time series collected](image1)

![Fig. 2 autocorrelation function of network traffic time series](image2)

![Fig. 3 partial correlation function of network traffic time series](image3)
Meanwhile it can be found from network traffic change curve chart that network traffic data sequence doesn't have certain seasonality and cyclic characteristics. Therefore, ARIMA and CAR non-stationary time series prediction models cannot be used for modeling and forecasting directly.

### 3.3 Network Traffic Wavelet Decomposition

For characteristics of network traffic non-stationary time series, wavelets are used for decomposition into stationary time series as described and then ARMA network traffic stationary time series are used for modeling and analysis as proposed in this paper. Db4 wavelet function in wavelet analysis is adopted for network traffic time series decomposition, and then restructuring operation is carried out through the low-pass reconstruction filter and high pass reconstruction filter. Network traffic time is broken into low-frequency components and high-frequency components, as detailed in Fig.4 and Fig.5.

#### Fig. 4 low-frequency components of network traffic

#### Fig. 5 high-frequency components of network traffic

### 3.4 Network Traffic Time Modeling

Firstly, low-frequency component A1 of network traffic data is used for stationarity testing and modeling. Autocorrelation function and partial correlation function of low-frequency components are shown in Fig.6 and Fig.7.

#### Fig. 6 autocorrelation function of network traffic low-frequency components
It indicates from Fig.6 that in Step 4 of time delay, the autocorrelation function of low-frequency component A1 starts to attenuate rapidly and it also indicates from Fig.7 that the partial correlation function of network traffic has trailing phenomena, suggesting that the low-frequency component A1 has become a stationary time series. ARMA model can be used for modeling and analysis. It indicates from autocorrelation function and partial correlation function of network traffic low-frequency components that using MA model for forecasting, exponent number of the model is 4. The least square method is used for identification of parameters of MA model and it is realized through DPS6.5 to get the following result:

\[ x_t = \varepsilon_t + 0.15\varepsilon_{t-1} - 0.774\varepsilon_{t-2} + 0.89\varepsilon_{t-3} - 0.165\varepsilon_{t-4} \]  

(4)

Where, \( \varepsilon_t \) stands for white noise sequence with average value of 0; \( x_t \) stands for low-frequency part of the network traffic time series.

Then high-frequency components D1 of network traffic data are used for stationarity testing and modeling. Autocorrelation function and partial correlation function of high-frequency components are shown in Fig.8 and Fig.9.

It indicates from Fig.8 that in Step 3 of time delay, the autocorrelation function of high-frequency component D1 starts to attenuate rapidly and the partial correlation function of network traffic has trailing phenomena, suggesting that the low-frequency component D1 has become a stationary time
series. ARMA model is used for modeling and analysis. It indicates from autocorrelation function and partial correlation function of network traffic high-frequency components that using AR model for forecasting, exponent number of the model is 3. The least square method is used for identification of parameters of AR model and it is realized through DPS8.5 to get the following result:

\[ x_t = \epsilon_t + 2.44\epsilon_{t-1} - 0.157\epsilon_{t-2} - 1.31\epsilon_{t-3} \]  

(5)

Where, \( \epsilon_t \) stands for white noise sequence with average value of 0; \( x_t \) stands for low-frequency part of the network traffic time series.

Combine network traffic values of low-frequency components and high-frequency components using the combinational formula \( S = A1 + D1 \) of the time series forecasting model, we can get the network traffic time series prediction model.

3.5 Results and Analysis

Fitting network traffic training samples according to the above model established and carrying out forecasting inspection for the test samples, we get the results as shown in Fig.10 and Fig.11. It is apparent from fitting result as shown in Fig.10 that the model has high fitting degree for the training sample, indicating that the model has sound fitting effect and the sample can be tested; It indicates from forecasting results as shown in Fig.11 that the accuracy of forecasting is quite high and forecasting results are satisfactory. Meanwhile it is compared with results of ARMA model without wavelet analysis. Comparison results show that network traffic models based on wavelet decomposition have high forecasting accuracy than which using the comparison models.

Wavelet analysis is a kind of nonlinear analysis method which is very suitable for the treatment of non-stationary data. It indicates from network traffic forecasting results that for non-stationary network traffic time series, through wavelet decomposition, non-stationary network traffic time series can be decomposed into single and stationary time series components, then we use stationary time analysis method to carry out modeling for stationary time series, which is favorable for network traffic time series modeling and it has further improved network traffic forecasting accuracy.

Fig. 10 network traffic time series fitting result
Meanwhile during practical application of network traffic, there are many factors leading non-stationary network traffic, such as cyclical factors and trend factor etc. Using a layer of wavelets for decomposition, it is difficult to get stable time series components. Therefore, multilayer decomposition is needed. But as the largest number of layers for decomposition in the non-stationary time series is strongly affected by the length of time series, when the length is short the largest number of layers for decomposition cannot meet the requirement of decomposing all layers. Thus dyadic wavelet conversion should be adopted for treatment, which should be noted during the practice of network traffic forecasting.

4. Conclusion

Network traffic forecasting is the basis of network management. As it is affected by economy, network users, and holidays and weekends etc., it has non-stationarity. It is difficult to obtain high accuracy forecasting results using the conventional time series analysis method. Therefore, a network traffic time series modeling method is proposed in this paper. Simulation experiments are carried out using specific network traffic data. Test results shown that comparing with forecasting results of the conventional time series forecasting method, using the time series forecasting method based on wavelet decomposition can improve the accuracy of network traffic forecasting and the results are satisfactory. It is a kind of stable and effective network traffic forecasting method.

5. References


