

Short Term Prediction Models of Mobile Network Traffic Based on Time Series Analysis

Yunxue Gao¹, Liming Zheng¹, Donglai Zhao¹, Yue Wu^{2(✉)},
and Gang Wang¹

¹ School of Electronic Information Engineering, Harbin Institute of Technology,
Harbin 150028, China

gaoyunxue@hotmail.com, {zheng, gwang51}@hit.edu.cn,
zd1527@126.com

² School of Management, Harbin Institute of Technology, Harbin 150028, China
wuy@hit.edu.cn

Abstract. In the mobile network, building a prediction based network traffic model is of great significance for mobile network optimization, so that the operators is able to schedule the resources adaptively. In the paper, multiplicative seasonal Autoregressive Integrated Moving Average model (ARIMA) and Holt-Winters model are proposed for modeling of traffic prediction, where the historical traffic series of a typical tourist area are utilized to verify the performance. The two methods analyze the trend of mobile network traffic per hour, build and validate models. Then predict mobile network traffic within a given period of time. The error rate of different models predictions is analyzed to provide certain decision basis for the allocation of network resources.

Keywords: Traffic model · Multiplicative seasonal ARIMA · Holt-Winters
Short term prediction

1 Introduction

With the continuous development and popularization of mobile communication technology and intelligent terminals, the demand of extracting dynamic information whenever and wherever possible is increasing. Keeping wireless networks with good and stable performance and reducing the occurrence of wireless network failures are becoming more important. Massive user groups increase the difficulty of operation and maintenance, and operators must allocate network resources reasonably to ensure the user experience. Low network load settings in hot spots may lead to increased congestion, decreased call quality and even paralysis of the system. Excessive network load will waste network resources. Therefore, the prediction of traffic in hot spots can provide a reasonable basis for decision-making and optimize the allocation of network resources.

Through the signaling analysis platform of both radio access network (RAN) and the core network that carriers deployed to collect mobile network signaling. The network and terminal data is intercepted from the signaling, and the data are cleaned and structured, and finally the data is acquired as time series. Researches on time series and

prediction have been discussed in [1]. In [2] authors proposed ARIMA model to process historical data of cells in mobile networks and generate reliable forecasting results. Traffic characterization is analyzed in [3] to be used to build models. In [4] authors proposed traffic prediction technique relies on analysis of traffic data on cells by using Holt-Winter’s exponential smoothing. The traffic prediction accuracy of machine learning techniques - Multi-Layer Perception (MLP), Multi-Layer Perception with Weight Decay (MLPWD) and Support Vector Machines (SVM) are investigated in [5]. A hybrid traffic prediction model is introduced to forecast the workload of base stations by utilizing historic traffic traces in [6]. In this paper, Holt-Winters and multiplicative seasonal ARIMA are used to model and validate the user data, and then traffic in the next period is predicted. The two models are popular models based on mathematic theory for time series short-term prediction. This article takes a typical tourist area as an example to build models and verify the performance.

2 Research Models

Time series is a sequence of data in equal intervals which mainly include long-term trend, seasonal variation and irregular change. The main purpose of time series analysis is to predict the future based on existing historical data. This section explains the principle of ARIMA and Holt-Winters.

2.1 ARIMA Model

ARIMA model doesn’t consider the influence of other factors. It only explores the change rule of the sequence itself. The model can be conducted to predict short-term and long-term data. But the precision of short-term forecast is higher [7]. The ARMA formula is:

$$\phi(B)\nabla^d x_t = \theta(B)e_t, \tag{1}$$

where $\{x_t\}$ is the observed time series, $\{e_t\}$ is a white noise sequence that satisfies a mean of zero and constant variance, B is the backward shift operator, and d is order of the difference. $\nabla = 1 - B$. $\phi(B) = 1 - \sum_{i=1}^p \phi_i B^i$, $\theta(B) = 1 - \sum_{j=1}^q \theta_j B^j$ where p and q is the order of AR and MA model, $\{\phi_j\}$, $\{\theta_j\}$ is the coefficients of AR and MA model.

The seasonality of time series is a common pattern that repeat over S time period. Multiplicative seasonal ARIMA is designed to fit model that has trends, seasonal characteristics and adjacent sequence correlation. Multiplicative seasonal ARIMA $ARIMA(p, d, q) \times (P, D, Q)_S$ includes seasonal and non-seasonal factors where P , D , Q are order of seasonal AR, seasonal difference and seasonal MA respectively. The formula is:

$$\phi(B)\Phi(B^S)\nabla^d \nabla_s^D x_t = \theta(B)\Theta(B^S)e_t, \tag{2}$$

where $\Phi(B) = 1 - \sum_{i=1}^P \Phi_i B^{iS}$, $\Theta(B) = 1 - \sum_{j=1}^Q \Theta_j B^{jS}$ and $\{\Phi_j\}$, $\{\Theta_j\}$ is the coefficients of AR and MA model.

2.2 Holt-Winters Model

Seasonal Holt-Winters smoothing fits data with trends and seasonality. The model can be applied to linear, exponential and damped trend. Holt-Winters includes the additive seasonal model and the multiplicative seasonal model. The additive seasonal model implies that seasonality and other trends are additive relation. Assume that α , β , $\gamma \in [0, 1]$ are smoothing parameters, a_t is the smoothed level, b_t is the trend, c_t is the seasonal smooth, and x'_{t+k} is k steps ahead forecasted value. The formulas [8, 9] are:

$$\begin{aligned} a_t &= \alpha(x_t - c_{t-s}) + (1 - \alpha)(a_{t-1} + b_{t-1}) \\ b_t &= \beta(a_t - a_{t-1}) + (1 - \beta)b_{t-1} \\ c_t &= \gamma(y_t - a_t) + (1 - \gamma)c_{t-s} \\ x'_{t+k} &= a_t + b_t k + c_{t+1+(k-1)\text{mod}S} \end{aligned} \quad (3)$$

The multiplicative seasonal model is used to fit data have multiplicative seasonality. The formulas [8, 9] are:

$$\begin{aligned} a_t &= \alpha(x_t/c_{t-s}) + (1 - \alpha)(a_{t-1} + b_{t-1}) \\ b_t &= \beta(a_t - a_{t-1}) + (1 - \beta)b_{t-1} \\ c_t &= \gamma(y_t/a_t) + (1 - \gamma)c_{t-s} \\ x'_{t+k} &= (a_t + b_t k)c_{t+1+(k-1)\text{mod}S} \end{aligned} \quad (4)$$

Holt-Winters is based on recursive relations, which suggests that start values need to be set before using. But start values have little effect on future calculations. For seasonal models, start values are detected by performing a simple decomposition in trend and seasonal component using moving averages.

3 Research Results

This section selects traffic data from 0:00 July 15, 2015 to 8:00 August 14, 2015 per hour as modeling data to explain the process of modeling. Forecast traffic data between 9:00 and 17:00 August 14, 2015. The experimental environment is R language with the version of 3.3.3.

3.1 ARIMA Modeling

The first step of model identification is to study the stationarity of time series. Stationary time series usually have a random fluctuation near a constant value, and the range of fluctuation is bounded with no obvious trend or seasonal characteristics. Non-stationary sequences tend to have different mean values at different time periods. The autocorrelation of stationary sequences decrease rapidly. The paper uses ADF (Augmented Dickey-Fuller Unit Root Test) to confirm whether time series are stationary. Null hypothesis forecasts the existence of unit root. If P value is smaller than 5%, the model test result reject the null hypothesis, and it can be assumed that time series satisfy stationarity. The function `adf.test` in R package `tseries` is used to run the

ADF test. The p value is less than 0.01, so the null hypothesis is rejected and the sequence is stable.

The second step is to determine the order of the model. Model order estimation is not only considering the fitting degree of the model to the original data, but also taking the number of undetermined parameters in the model into account, and take a reasonable tradeoff between the two. AIC (Akaike information criterion) is applicable to the order determination. Try to choose the minimum AIC value within a certain range. And Try to avoid over fitting as much as possible. Several excellent models and their AIC values are shown in Table 1. According to the simulation results, the Θ_2 parameter of $ARIMA(1, 0, 2) \times (1, 0, 2)_{24}$ is 0.0680 approximately zero, the t value is not significant, and the AIC value is close to $ARIMA(1, 0, 2) \times (1, 0, 1)_{24}$. In order to simplify the complexity, $ARIMA(1, 0, 2) \times (1, 0, 1)_{24}$ is applied. Model identification is completed.

Table 1. Several models and their AIC values

Model	AIC
ARIMA(1, 0, 2)(1, 0, 2)[24]	4107.25
ARIMA(1, 0, 2)(1, 0, 1)[24]	4108.32
ARIMA(2, 0, 1)(1, 0, 1)[24]	4109.03
ARIMA(1, 0, 1)(1, 0, 2)[24]	4112.04

Using MLE (Maximum Likelihood Estimation) to estimate parameters of the model. These parameters are shown in Table 2.

Table 2. Parameters of $ARIMA(1, 0, 2) \times (1, 0, 1)_{24}$

ϕ_1	θ_1	θ_2	ϕ_1	Θ_1	Intercept
0.7424	0.5933	0.1113	1	-0.9854	21.4832

After the completion of model identification, order determination and parameter estimation, next step is to estimate the validity of the chosen model. The validity test of the model is essentially to examine whether the residual sequence is a white noise sequence. The time series diagram of residuals should be a rectangular scatter plot without any trend around zero. The mean value of residuals is calculated to be about 0.09. And residual sequence meets the requirements according to Fig. 1. The auto-correlation function (ACF) of residual sequence is calculated. The correlation coefficients of the residuals are small except the zeroth order in Fig. 1, which means the residuals satisfy the independence. ACF only considers the existence of a certain lag order. The Ljung-Box test is based on a series of lags to determine the correlation of the sequence. The p values in Fig. 1 are all greater than 0.05, so the null hypothesis cannot be rejected, which means the residual sequence is a white noise sequence.

Once the model is identified, future values can be predicted from past and present values of the time series. Minimum mean square error estimation is commonly used in prediction. Confidence intervals, predictive values and actual values are displayed in Fig. 2.

3.2 Holt-Winters Modeling

The additive Holt-Winters model is used to analysis data. This section analysis and predict the same period of time series as previous section.

The model uses following parameters: 0.102 for a, 0.004 for b, and 0.231 for c. The value of B is approximately 0, indicating that the slope of the estimated trend component is essentially constant. The value of ‘a’ and ‘c’ indicate that the level and seasonality based on recent observations and historical observations. The ACF of residual sequence is calculated in Fig. 2. Based on the additive Holt-Winters model, the prediction of time series is calculated and displayed in Fig. 3.

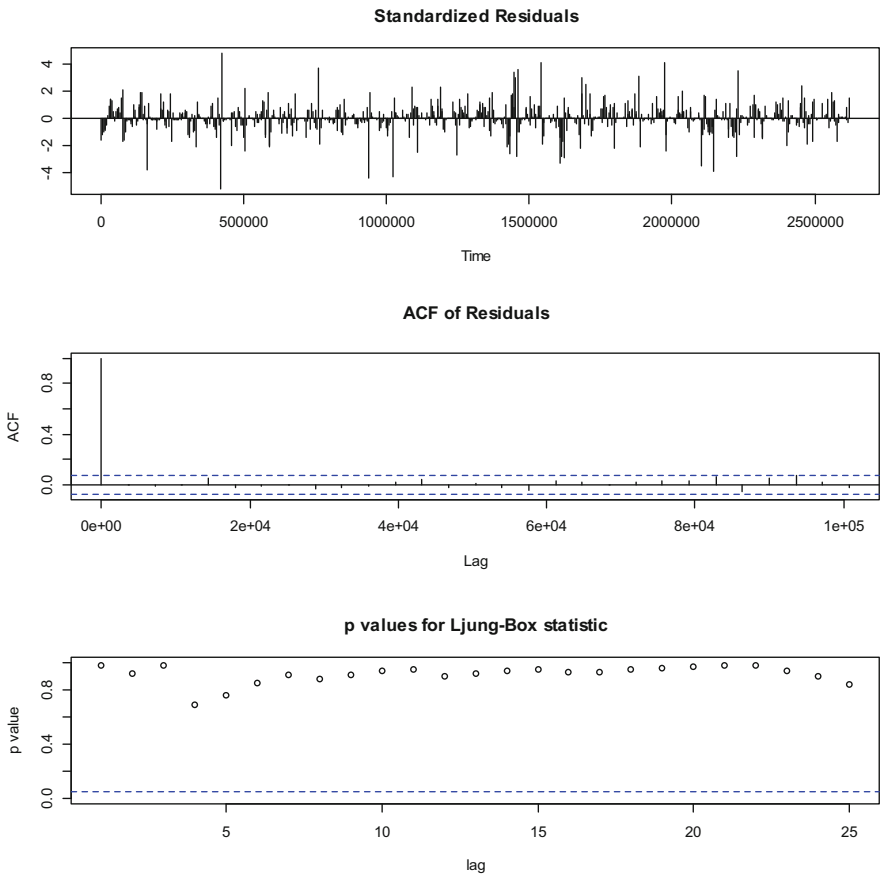


Fig. 1. This shows three figures consisting of *standardized residuals* (first), *ACF of residuals* (second), and *p value for Ljung-Box statistic* (third).

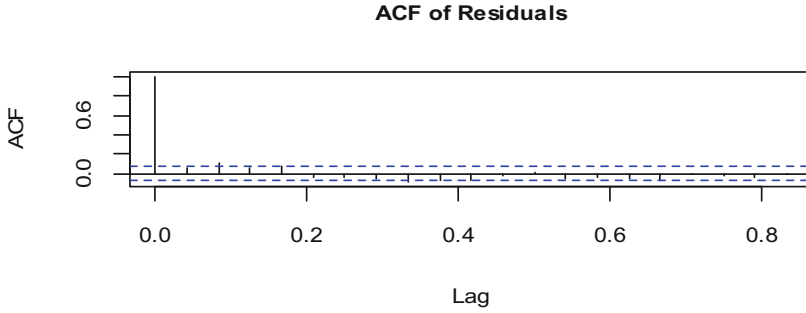


Fig. 2. ACF of residuals

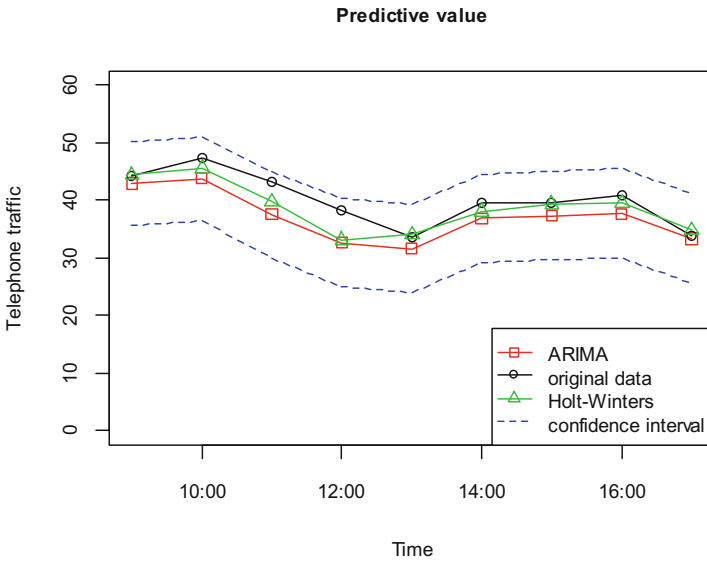


Fig. 3. Predictive value of ARIMA and Holt-Winters

4 Conclusion

Accurate traffic prediction is beneficial to balance the resources of mobile network according to the different characteristics of busy and idle time. Prediction helps to avoid excessive network load or waste of network resources. Depending on the data structure, different approaches can be tried. The paper presented the analysis of mobile network traffic. Multiplicative seasonal ARIMA model and Holt-Winters model are provided. The two models build and validate models and predict future values respectively. The forecast error of Holt-Winters model is lower than ARIMA in this situation. The order determination of ARIMA is more complex. Holt-Winters model is better than

multiplicative seasonal ARIMA model in this situation. However, the model adopted in reality is decided by the different characteristics of the specific data, and a model can not always be the optimal model.

Acknowledgement. This work was supported by the National Natural Science Foundation of China under Grant No. 61401120.

References

1. Brockwell, P.J., Davis, R.A.: Introduction to Time Series and Forecasting. Springer, New York (2002). <https://doi.org/10.1007/b97391>
2. Guo, J., Peng, Y., Peng, X.: Traffic forecasting for mobile networks with multiplicative seasonal ARIMA models. In: 2009 9th International Conference on Electronic Measurement & Instruments (2009)
3. Khedher, H., Valois, F.: Traffic characterization for mobile networks. In: IEEE 56th Vehicular Technology Conference, pp. 1485–1489 (2008)
4. Tikunov, D., Nishimura, T.: Traffic prediction for mobile network using Holt-Winter's exponential smoothing. In: 15th International Conference on Software, Telecommunication and Computer Networks, pp. 310–314 (2007)
5. Nikravesh, A.Y., Ajila, S.A., Lung, C.-H.: Mobile network traffic prediction using MLP, MLPWD, and SVM. In: BigData Congress (2016)
6. Dawoud, S., Uzun, A., Göndör, S.: Optimizing the power consumption of mobile networks based on traffic prediction. In: Computer Software and Applications Conference (COMPSAC) (2014)
7. Mladenović, J., Lepojević, V., Janković-Milić, V.: Modelling and prognosis of the export of the Republic of Serbia by using seasonal Holt-Winters and ARIMA method. *J. Econ. Themes* **54**, 233–260 (2016)
8. Koehler, A., Snyder, R., Ord, J.K.: Forecasting models and prediction intervals for the multiplicative Holt-Winters method. *Int. J. Forecast.* **17**, 269–286 (2001)
9. Lepojevic, V., Andjelkovic-Pesic, M.: Forecasting electricity consumption by using Holt-Winters and seasonal regression models. *Econ. Organ.* **8**(4), 421–431 (2011). *Facta Universitatis, Series*