A Traffic Interval Prediction Method Based on ARIMA

Tiantian Lv*, Yanqin Wu and Le Zhang
Research Institute of China Telecom Corporation Limited, Beijing, China
lvtt@chinatelecom.cn

Abstract. The purpose of this paper is to construct an autoregressive integrated moving average model (ARIMA) for forecasting the traffic interval, which is helpful for the mobile industry to forecast the change of requirements for the peak of customer traffic, adjust the bandwidth dynamically, and improve the ability of active service. In this paper, the lateral time series analysis method is applied to analyze the data of the peak traffic of the uplink and downlink network from August to October of 2017 and 2018 to establish ARIMA prediction model for determining the parameters of it. MAPE method is used for model assessment and model diagnosis. Then, the optimal forecasting model is selected and the forecast error rate is calculated as the adjustment parameter of the forecast range. The model is used to forecast the customer flow range in the first three days of October 2019. Finally, the MIA method is proposed to compare with the LSTM algorithm and linear regression method for interval cumulative error comparison of traffic interval prediction. The result shows that the ARIMA (0,1,0) model has the lowest mean interval error rate of 15.5%, which proves the reliability of the model in predicting peak traffic interval.

1. Introduction
With the rapid development of internet technology, the explosive growth of network users and the richness of network service types have greatly increased the complexity of network [1]. Operators usually set the bandwidth of the network according to the size of the historical data, which cannot dynamically adjust the bandwidth to meet the requirements of customers. The traffic prediction of network has important meaning in network control and traffic management, which is helpful to network planning, and better dynamic planning of various network resources, such as dynamic bandwidth adjustment, to meet the requirements of customers and cut the cost at the same time [2].

2. Research status
There are lots of research about network traffic prediction. Hu Sheng[3] proposed a method of constructing network traffic model based on ARIMA, which can make network traffic prediction for Measured traffic data based on ARIMA model, and have high accuracy rate for analysis and prediction of partial nonstationary network traffic data. Zhongda Tian[4] proposed a network traffic forecasting model based on gauss process regression compensating ARIMA, which uses ARIMA model to predict the linear component of network traffic, and uses gauss process regression model to predict the error of network traffic prediction with nonlinear characteristics. Ran Zhang[5] proposed application research of ARIMA Model in network traffic prediction, which uses the difference method to smooth the original data of network traffic, establishes the optimal prediction model of network traffic, and finally simulates and forecasts the data of network traffic. Manuel Lopez-Martin[6] proposed neural network architecture based on gradient boosting for IoT traffic prediction, which uses stochastic
gradient descent to train the resulting architecture end to end, uses real data on IoT traffic from mobile operators to predict network traffic. Michal Aibin[7] proposed a traffic prediction based on machine learning for elastic optical networks, which uses machine learning technology to realize traffic prediction in elastic optical networks.

The above methods only consider network traffic prediction for single characteristic. But the current network traffic is a nonlinear, multiple time dimensions dynamic system with self-correlation and burst characteristics, which is necessary to predict the traffic according to the data characteristics of network traffic to improve the accuracy of prediction. At the same time, point prediction can not reflect the randomness and uncertainty of the data set, but predicting an interval with a certain probability can solve this problem.

Through the analysis of the above problems, this paper proposes a traffic interval prediction method based on ARIMA, which, based on the interval prediction, can better reflect the actual network customer future demand daily change trend as well as the flow peak value interval scope, judge the future time flow whether to exceed the limit, determine whether to adjust the bandwidth rate at which an existing circuit is actually ordered. This method also indirectly indicates the necessity of interval prediction.

3. Related theories and methods

3.1. ARIMA model

The full name of ARIMA model is autoregressive moving average model, which can be used to forecast time series and is often used in demand forecasting and planning. There are three parameters of the ARIMA model: p, d, and q, and it is usually written as an ARIMA (p, d, q) model. Among them, “AR” is “autoregressive”, “p” is the number of autoregressive terms, “MA” is “moving average”, “d” is the number of differences made to make it a stationary sequence, and “q” is the lag of the prediction error used in the prediction model Number, which is also known as the number of moving average terms.

The mathematical formula of ARIMA is:

\[ \Delta^d y_t = \mu + \sum_{i=1}^{p} \varphi_i \Delta^d y_{t-i} + \sum_{j=1}^{q} \theta_j \Delta^d e_{t-j} \]  

Where \( \mu \) is the mean of the original time series, \( y_t \) is the original time series, \( \Delta^d y_t \) represents the stationary sequence of \( y_t \) after \( d \) differences, \( e_t \) represents the white noise random error sequence with a mean value of 0, \( \varphi \) represents the coefficient of AR, and \( \theta \) represents the coefficient of MA.

3.2. Traffic interval prediction method based on ARIMA

To construct an ARIMA model, first of all, the stationarity of time series data needs to be judged. Then the non-stationary time series need to be differentiated, and the parameters p and q are determined after reaching the plateau. The traffic interval prediction model based on ARIMA is added error rate calculation and prediction interval calculation based on traditional ARIMA model. The specific traffic interval prediction method based on ARIMA is as follows:

1) Data pre-processing: including the filling of missing values, the processing of outliers and the unification of units.

2) Stationary analysis: test the stationarity of time series data.

3) Difference calculation: if the time series data is nonstationary, different calculation is needed until it is stationary.

4) Autocorrelation coefficient: analyse the autocorrelation coefficient figure (ACF) and partial autocorrelation coefficient figure (PACF) of the stationary time series after the difference processing, determine the values of the parameters p and q, and establish the ARIMA model.

5) Model checking: use the ARIMA model to predict the network traffic data in September 2017, use the MAPE method to check the model, and adjust the model parameters.
6) Deviation rate calculation of interval: based on the tested ARIMA model the inflow and outflow for the first three days of October 2017 and 2018 are predicted respectively using the inflow data and the outflow data in September 2017 and September 2018. The predicted value is calculated by the difference inverse operation and compared with the actual value, the average error rate for inflow and outflow are calculated respectively, named sin and sout respectively, to be used as deviation rate calculation of the predicted interval.

7) Traffic interval prediction: inflow and outflow in September 2019 are used to predict the inflow and outflow in the first days of October 2020 respectively to obtain predicted values for inflow, named pin, and outflow, named pout, and the predicted interval of inflow is \([pin \ast (1 - sin), pin \ast (1 + Sin)]\), and the predicted interval of outflow is \([pout \ast (1 - sout), pin \ast (1 + sout)]\).

8) Inverse difference calculation: if the time series data is nonstationary, inverse different calculation is needed to make the difference data to be restored.

9) Error rate calculation of interval: finally, the MIE method in formula 3 is used to calculate the error rate of the predicted traffic interval.

The process of constructing ARIMA model and predicting traffic interval based on ARIMA in this paper is shown in Figure 1.

**Figure 1.** The process of traffic interval prediction method based on ARIMA

### 4. The evaluation method of model

#### 4.1. Model parameter evaluation method of ARIMA

General point prediction evaluation methods include Mean Absolute Error Method (MAE), Mean Square Error Method (MSE), Root Mean Square Error Method (RMSE), Mean Absolute Percent Error Method (MAPE), etc. This document is based on MAPE method for ARIMA model parameters Adjustment, which calculation formula is:

\[
MAPE = \frac{\sum_{i=1}^{n} \left| \frac{Y_i - F_i}{Y_i} \right| \times 100\%}{n}
\]

Where \(Y_i\) is the actual value, \(F_i\) is the predicted value, and \(n\) is the number of predicted data.

#### 4.2. Evaluation method of error rate for traffic interval prediction

The positional relationship between the actual value and the predicted value needs to be considered to calculate the error rate of the traffic interval prediction. This document proposes a method of the average cumulative error rate to predict traffic interval, which can calculate the error rate of the traffic prediction interval through the relationship between the actual value and the boundary value of the traffic prediction interval.
When the actual value is not within the predicted interval, the smaller the deviation between the actual value and the predicted interval, the better the model effect. To illustrate the extent of the shift, the mean interval error rate (MIE) at the i-th time point is defined as follows:

\[ MIE = \frac{1}{m} \sum_{i=1}^{m} \epsilon_i \]  

\[ \epsilon_i = \begin{cases} \frac{L(F_i) - Y_i}{Y_i} \times 100\% , & \text{if } Y_i < L(F_i) \\ 0 \% , & \text{if } Y_i \in [L(F_i), H(F_i)] \\ \frac{Y_i - U(F_i)}{Y_i} \times 100\% , & \text{if } Y_i > U(F_i) \end{cases} \]  

Among them, \( L(F_i) \) is the minimum value of the prediction interval, \( U(F_i) \) is the maximum value of the prediction interval, \( Y_i \) is the actual value, and \( m \) is the number of days in the prediction interval.

5. Experimental results and analysis

5.1. Network traffic data pre-processing

In the process of network traffic data collection, there are some missing for them, which is necessary to be filled before been predicted. There are multiple methods for filling missing data, such as single filling method, mean filling method, median filling method and k nearest neighbor method and so on [8].

Because only a few random values are missing in the peak data of network traffic, the total missing rate is less than 10\%, so the median filling method is used in this document to deal with the missing values.

5.2. Parameter analysis of ARIMA model

In this document, data of September 2017 and September 2018 are used to predict inflow and outflow for the first days of October, respectively. First of all, the stability test of data is performed with SPSS tool. The results of the traffic smoothness analysis for September 2017 are shown in Figure 2.

![Figure 2](image.png)

**Figure 2.** Results of the traffic smoothness analysis for September 2017

Figure 2 shows that the inflow and outflow sequence of September 2017 are both non-stationary, so the different operation of the sequence is needed. The figure of sequence after the first order difference is shown in Figure 3.
Figure 3. Result of first order difference

It can be seen from Figure 3 that the series after first order different oscillates uniformly above and below 0, which is known that the series accords with the stationarity, so the parameter D in ARIMA model is 1. The ACF and PACF test results are performed below and the results are as shown in Figure 4.

Figure 4. ACF and PACF test results

ACF and PACF test results shows that the order of confidence does not exceed the confidence interval, so the value of p and q in the ARIMA model for inflow and outflow are both 0, and the
ARIMA model of inflow and outflow are both ARIMA (0,1,0). Since 2018 and 2019 traffic data have the same stationarity characteristics as 2017, the same model is used for traffic prediction of them.

5.3 ARIMA model test

Because the ARIMA model parameters determined by ACF and PACF may have error or one-sidedness, which may make ARIMA model based on global data is not necessarily optimal. Therefore, formula (2) is used to compare the ARIMA models with different parameters. The MAPE values of ARIMA (0,1,0), ARIMA (1,1,0), ARIMA (0,1,1) and ARIMA (0,2,0) are compared using the same experimental data used in section 5.2. The results are as follows:

| Model        | MAPE-value |
|--------------|------------|
| ARIMA (0,1,0)| 8.3%       |
| ARIMA (1,1,0)| 16.5%      |
| ARIMA (0,1,1)| 26.7%      |
| ARIMA (0,2,0)| 14.8%      |

It can be seen from the Table 1 that MAPE value of ARIMA (0,1,0) model is the lowest, which shows the reliability of parameter selection of this model.

5.4 Evaluation of traffic interval prediction based on ARIMA

In order to prove the reliability of the method of ARIMA (0,1,0), formula (3) is used in this section to compare the accuracy between ARIMA (0,1,0) model, LSTM model and linear regression model. The comparison is shown in Table 2.

| Model               | Mean interval error rate of inflow | Mean interval error rate of outflow | Mean interval error rate of inflow and outflow |
|---------------------|------------------------------------|------------------------------------|-----------------------------------------------|
| ARIMA (0,1,0)       | 17.5%                              | 13.5%                              | 15.5%                                         |
| LSTM                | 39.5%                              | 25.6%                              | 27.7%                                         |
| linear regression   | 48%                                | 58.4%                              | 46.8%                                         |

Table 2 shows that the mean interval error rate of inflow and outflow based on ARIMA (0,1,0) model is 15.5%, lower than the results of LSTM model and linear regression model, which shows the reliability of ARIMA (0,1,0) model.

6. Conclusion

In this document, a traffic interval prediction method based on ARIMA is proposed. The MAPE method is used to choose the parameters of ARIMA model. After comparison, ARIMA (0,1,0) model is chosen to calculate the traffic interval of network inflow and outflow. MIE method is proposed in this document to compare the mean interval error rate of ARIMA (0,1,0) model, LSTM model and linear regression model. Finally, the result of the experiment evaluation shows that the mean error rate of predicting interval of ARIMA (0,1,0) model is lower than LSTM model and linear regression model, which proves the reliability of the ARIMA (0,1,0) model.

Acknowledgments

The authors of this document would like to thank assistance and encouragement from Yanqin Wu and Le Zhang. The authors also thank financial support from China SASAC project-SDN collaborative controller 20HQBYYF0093-001.
References

[1] ZHAO J L, QU H, ZHAO J H, et al. A comprehensive forecasting model for network traffic based on Morlet-SVR and ARIMA[J]. Journal of Beijing University of Posts and Telecommunications, 2016, 39(2): 5357.

[2] DUAN H Q, TANG B H. Prediction of data flow in computer network based on linear multiscale model[J]. Journal of Shenyang University of Technology, 2017, 39(3): 322-327.

[3] SHENG Hu, ZHANG Y X. Network Traffic Modeling and Forecasting based on ARIMA[J]. Communications Technology, 2019, 52(04): 903-907.

[4] Tian Z D, Li S J, Wang Y H, Wang X D. Network Traffic Prediction Based on ARIMA with Gaussian Process Regression Compensation[J]. Journal of Beijing University of Posts and Telecommunications, 2017, 40(06): 65-73.

[5] Zhang L, Zhao C L. Application Research on Network Traffic Prediction Base on ARIMA[J]. Computer Simulation, 2011, 28(02): 171-174.

[6] Manuel LM, Belen C, Antonio SE. Neural network architecture based on gradient boosting for IoT traffic prediction. Future Generation Computer Systems, Volume 100, 2019, Pages 656-673, ISSN 0167-739X.

[7] Michal A. Traffic prediction based on machine learning for elastic optical networks, Optical Switching and Networking, Volume 30, 2018, Pages 33-39, ISSN 1573-4277.

[8] Liu Z Y. Data cleaning technology and application based on big data [J]. Data Technology and Application, 2019, 37(04): 92-94.