A multi-applications comprehensive traffic prediction model for the electric power data network

Yu Zhou
State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications,
Beijing, China
yzhou@bupt.edu.cn

Yutong Ji
State Grid Jibei Electric Power Company Limited Information & Communication Dispatch,
Beijing, China
15901016220@163.com

Ningzhe Xing, *the corresponding author*
State Grid Jibei Electric Power Company Limited Information & Communication Dispatch,
Beijing, China
xingningzhe@163.com

Wenjing Li and Shao-Yong Guo
State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications,
Beijing, China
{wjli, syguo} @bupt.edu.cn

***Abstract*— Currently, the requirements of service quality in the electric power data network are getting higher and higher, and traffic prediction is an important premise to promote service quality. In order to accurately predict the total traffic of communication channels, a Multi-Applications Comprehensive Traffic Prediction (MACTP) model is proposed in this paper. Differing from F-ARIMA and S-ARIMA models which are used to predict the traffic of single application, the proposed MACTP model is used to predict the traffic of multi-applications conveyed in the channels. Simulation results show that MACTP model has higher accuracy and efficiency than classical prediction models, and it is suitable for electrical power data network.**

***Keywords*— traffic prediction; electrial power data network; multi-applications;F-ARIMA model; S-ARIMA model**

I. INTRODUCTION

With the construction of strong smart grid, the electric power data network and its supporting application systems have been developed rapidly. Requirements of service quality in the electric power data network are also getting higher and higher. Optimally allocating the network bandwidth is an effective approach to improve service quality. And traffic prediction is the premise of bandwidth allocation. So it is important to predict the practical traffic condition accurately to ensure the quality of the electric power data network and avoid network congestion.

At the beginning, Fourier series model is widely used in traffic prediction [1]. However, due to the randomness of traffic data, Fourier series model is limited by processing random data. In the current papers, Holt-winter 's model has better performance when the traffic is high (over 5Erlang), but the performance is not good when the traffic is low (below 5Erlang) [2]. But most of the actual traffic volume is not in the numerical range that Holt-winter 's model acts well in. Besides, the enlargement of the data's randomness has a great impact on the precision of Holt-winter 's model. As one of time series models, ARIMA model has been used to resolve above problems. But the model changes greatly on account of different sample points. At the same time, the prediction steps is short [3]. Besides, F-ARIMA model (Fractional-ARIMA model) and S-ARIMA model (Seasonal-ARIMA) are the extension of ARIMA model [4-5].

Currently all model systems are mainly used to model and predict single type of traffic. However, the traffic need to be predicted from a macro perspective of the entire channel that supports various applications, because the fluctuations of individual business traffic can neither bring network hidden danger, nor cause network congestion. Therefore, according to the practical situation of the electric power data network, the paper puts forward a Multi-Applications Comprehensive Traffic Prediction (MACTP) model based on F-ARIMA and S-ARIMA models. Because of considering the flow characteristics of different application types, the prediction effects of the model are more accurate and effective.

The remainder of this paper is organized as follows. Section II divides applications into three types: voice, video and data, and analyzes their characteristics. Section III proposes MACPT model on the basis of application characteristics. Section IV presents the simulation and comparison analysis. Finally, conclusions are drawn and the future work is discussed in Section V.

II. CHARACTERISTICS ANALYSIS OF APPLICATIONS

At present, application types of the electric power data network can be classified into voice, video and data. Traffic characteristics of these application types are analyzed in the following.

1.Voice application: voice application is the traditional application in the electric power data network, mainly including dispatching telephones, office telephones and conference calls, etc.. The requirements of reliability and real-time for voice application are relatively high, but the bandwidth requirement is not high.

2.Video application: video application is one of indispensable basic application in the multimedia monitoring field, mainly including video conferencing and unattended substations and other video surveillance. The characteristics of this kind of application is that requirements of network latency and bandwidth are higher.

3.Data application: it mainly includes SCADA (Supervisory Control And Data Acquisition) data application and MIS (Management Information System) data application [6]. SCADA data application requires high reliability and high real-time, and its traffic is generally small (300Kbps to 800Kbps), besides, the delay requirement is higher. MIS data stream can bring periodic bursts of traffic (peak to 4~6Mbps), so demand for the network transmission bandwidth is high, and there is no strict requirements on the network delay.

Application traffic not only has a short-range dependence, but presents the self-similarity in a larger time scale, which can also be called a long-range dependence. F-ARIMA model is used to process video traffic against its long-range dependence and S-ARIMA model is used to fit voice and data traffic aimed at the burst and periodicity. On basis of the aforesaid discussion, MACTP model is put forward in this paper.

III. MACTP MODEL

A. Multi-applications comprehensive traffic prediction model

F-ARIMA model and S-ARIMA model are introduced by references [4-5]. As for the burst of voice and data application and the long term of video application, we fit voice and data traffic data based on S-ARIMA model and fit video traffic based on F-ARIMA model. And then, the following MACTP model is proposed:

$$Z_t = \sum_{i=1}^{3} \varphi_i(B)\Psi_i(B^{s_i})\omega_i^{-1}(B)\Omega_i^{-1}(B^{s_i})\Delta^{-d_i}\Delta_{s_i}^{-D_i}\varepsilon_t \tag{1}$$

In the above equation,

$$\begin{cases} \Delta = 1 - B \\ \Delta_{s_i} = 1 - B^{s_i} \end{cases} \tag{2}$$

$$\begin{cases} \omega_i(B) = 1 - \sum_{k=1}^{g_i} \omega_{i,k} B^k \\ \Omega_i(B^{s_i}) = 1 - \sum_{k=1}^{G_i} \Omega_{i,k} B^{s_i k} \end{cases} \tag{3}$$

$$\begin{cases} \varphi_i(B) = 1 - \sum_{k=1}^{l_i} \varphi_{i,k} B^k \\ \Psi_i(B^{s_i}) = 1 - \sum_{k=1}^{L_i} \Psi_{i,k} B^{s_i k} \end{cases} \tag{4}$$

Especially, when $i = 2$,

$$\Omega_i(B^{s_i}) = \Psi_i(B^{s_i}) = \Delta_{s_i}^{-D_i} = 1$$

where: $i = 1,2,3$ is on behalf of the voice, video, data, three types of application respectively. B is the backward shift operator. $\{\varepsilon_t\}$ is a white noise sequence with zero-mean. $\omega_i(B)$ and $\varphi_i(B)$ are the conventional auto regressive operator and the conventional moving average operator (or polynomials).$\Omega_i(B^{s_i})$ and $\Psi_i(B^{s_i})$ are the seasonal auto regressive operator and the seasonal moving average operator (or polynomials).g_i, G_i, l_i, L_i are order of operators. d_i is the differential order. D_i is the seasonal differential order. Δ^{d_i} is the differential operator. $\Delta_{s_i}^{D_i}$ is the seasonal differential operator. s is seasonal periodicity.

B. Parameters setting

The following introduces methods of setting the related parameters:

1.The differential order d_i

The literature shows the relation of $d = H - 0.5$ between the differential order d_i and the Hurst index (H) of time series. Therefore, we can figure out Hurst index first, and then get the differential order d_i.We choose the R/S analysis method to calculate the Hurst index [7].

2.The model order g_i, l_i

We set the model order based on autocorrelation and partial autocorrelation function, meanwhile, BIC is used to verify the terseness.

1) A method based on autocorrelation and partial autocorrelation function. This method judges the model order by the truncation of autocorrelation and partial autocorrelation function $\{\rho_k\}$,$\{\varphi_{kk}\}$ [8]. The process of judging truncation is given detailedly in Section IV.

TABLE I. JUDGING TRUNCATION

	AR(g)	MA(l)	ARMA(g, l)
$\{\rho_k\}$	Tailing	Censoring	Tailing
$\{\varphi_{kk}\}$	Censoring	Tailing	Tailing

2) A method based on information criterion.

BIC (Bayesian Information Criterion): Its function is $BIC(g,l) = n\ln\hat{\sigma}^2(g,l) + (g+l)\ln n$. where; $\hat{\sigma}^2(g,l)$ is the variance of fitting model residuals. The upper limit of g,l generally takes $\sqrt{n}$ or $\frac{n}{10}$. The method is in order to obtain the minimum value of g and l.

3. The operator coefficient $\omega_{i,k}$, $\Omega_{i,k}$, $\varphi_{i,k}$, $\Psi_{i,k}$

The estimation methods of the operator coefficients are moment estimation, maximum likelihood estimation and least square method [9]. Maximum likelihood method and least square method make full use of the information of each sample value, which are accurate estimation. Maximum likelihood method is used to estimate the coefficients of each operator in this paper.

IV. MODEL SIMULATION ANALYSIS AND COMPARISON

A. Simulation process

The steps to build MACTP model are shown below:

Step 1: Data acquisition

A large number of probes are deployed in the power grid to collect information.

Step 2: Data preprocess

Next, the data should be preprocessed, and it is mainly divided into four steps:

1) Dividing data into voice, video and data types, based on the analysis of the protocol and the port.

2) Doing aggregation operation for these three types of application in a certain scale. A clustering method is to take data blocks of stationary time series, $X = \{X(i), i > 0\}$, at the length of 1 min or 1 h for the sum operation, that is $X_m(k) = \sum_{i=1}^{k} X_m(i)$, $m = 1,2,3 \ldots, k = 0,1,2 \ldots$

3) Dealing with the given network traffic data to get a time series with a mean value of zero.

4) Denoise of sequence. Common methods are the window function, the exponential filter, the wavelet transform, etc..

Step 3: The model order for three types of application traffic need to be determined. Selecting appropriate value of g, d, l, G, D, L to get a tentative model.

Step 4: A preliminary estimate of the parameters of MACTP model is obtained by the autocorrelation value of the pretreated sequence.

Step 5: Residual test. If the model is not ideal, then re-fitting and test again.

Step 6: Prediction and analyzation. Compare MACPT model with other prediction models by calculating the error.

B. Estimation indexes

We use ME (Mean Error), MAD (Mean Absolute Deviation), MSE (Mean Square Error), MPE (Mean Percentage Error) and MAPE (Mean Absolute Percentage Error) to judge the prediction accuracy. Specific calculation formula is as follows:

Mean Error:

$$\delta_{ME} = \frac{\sum_{i=1}^{n}(\hat{Z}_i - Z_i)}{n} \tag{5}$$

Mean Absolute Deviation:

$$\delta_{MAD} = \frac{\sum_{i=1}^{n}|\hat{Z}_i - Z_i|}{n} \tag{6}$$

Mean Square Error:

$$\delta_{MSE} = \frac{\sum_{i=1}^{n}(\hat{Z}_i - Z_i)^2}{n} \tag{7}$$

Mean Percentage Error:

$$\delta_{MPE} = \frac{\sum_{i=1}^{n}(\frac{\hat{Z}_i - Z_i}{Z_i} \times 100\%)}{n} \tag{8}$$

Mean Absolute Percentage Error:

$$\delta_{MAPE} = \frac{\sum_{i=1}^{n}(\frac{|\hat{Z}_i - Z_i|}{Z_i} \times 100\%)}{n} \tag{9}$$

where: $\hat{Z}_i$ is predictions of time series. Z_i is actual values of time series. n represents total steps. δ is on behalf of error values calculated by different methods. An error value is lower, prediction precision is higher.

C. Simulation results

We use RStudio and MATLAB to process data and emulate.

A zero mean sequence is a premise to establish MACTP model. After zero mean normalization and denoise operation to three application types of traffic, R/S method is used to calculate Hurst index of a time series and the differential order.

TABLE II. COMPARISON ABOUT H, D BEFORE AND AFTER DIFFERENCE

Application types	Before differential operation		After differential operation	
	H	d	H	d
Voice	0.8237	0.3237	0.5124	0.0124
Video	0.7776	0.2776	0.5093	0.0093
Data	0.8578	0.3578	0.5112	0.0112

In this paper, the autocorrelation and partial autocorrelation function is used to determine the model order. Fig.1 is a set of diagrams of autocorrelation and partial autocorrelation values. Among them, the first column is the ACF chart and the second column is the PACF chart. In the figure, the horizontal axis is

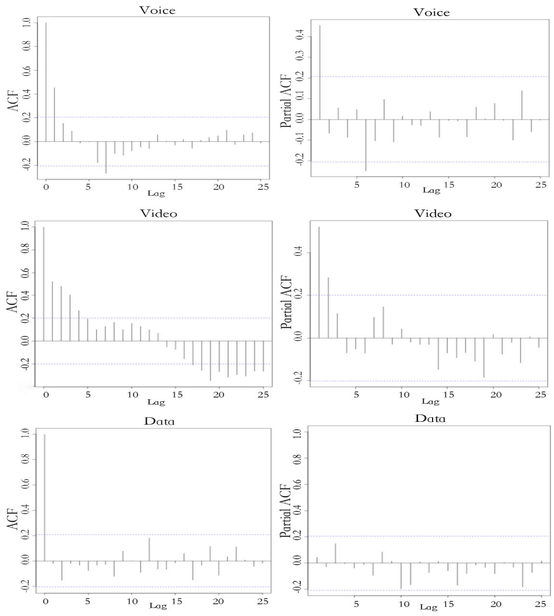

Fig. 1. The determination of orders

the order of delay and the vertical axis represents ACF and PACF values. Besides, the two blue horizontal lines in the figure indicate the 95% confidence interval.

As shown in Fig.1, for voice application, ACF rebounds at order 6 and reenters the confidence interval after order 7, then the whole tails, so we take l_1 as 6. PACF is beyond confidence interval in 6 steps obviously, so we take g_1 as 6. Similarly,for video application, we take l_2 as 6 and g_2 as 0. For data application, we take l_3 as 3 and g_3 as 3.

Taking into the periodicity of the autocorrelation and partial autocorrelation, this paper treats voice and data traffic with first-order seasonal differential operation. In addition, in the process of the actual simulation, it is found that the AR (4) and the later part of the model have little effect, so these parts are excluded, considering the simplicity of a model according to AIC and BIC. Finally, fitting models for voice, video and data application are determined as: $S - ARIMA(3,0.3,6) \times (0,1,1)_6$, $F - ARIMA(0,0.2,6)$ and $S - ARIMA(3,0.4.3) \times (0,1,1)_{24}$.

The predicted result is shown below.

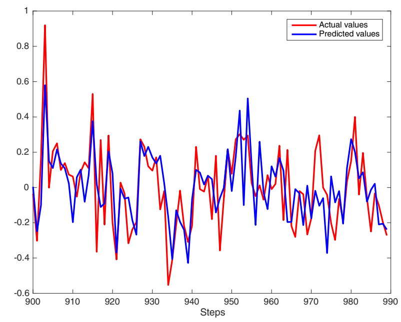

Fig. 2. Prediction results

The model reflects the trend of traffic well in Fig.2. The predictive value of the original flow can be obtained by the inverse operation of differential, zero mean normalization and the denoise to the prediction data.

D. Comparative analysis

Fourier series model, Holt - winter 's model, and MACTP model are used to predict traffic by total flow values within one day. And results are compared with actual flows. Estimation indexes are calculated and displayed in Fig.3.

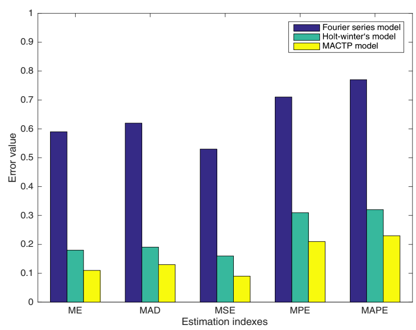

Fig. 3. Prediction accuracy comparison of three models

Fourier series model is used to process data with weak randomness and strong periodicity well, but actual flows within one day have strong randomness and weak periodicity, so the accuracy of Fourier series model is poor. Holt-winter 's model can adapt to the randomness better. But when traffic volume is low, prediction effect is bad.

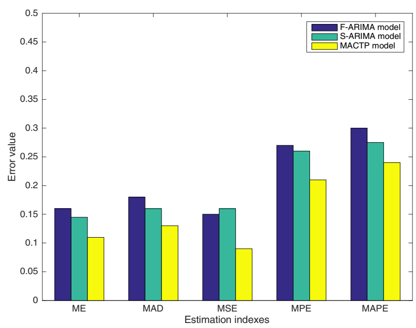

Fig. 4. Prediction accuracy of two models without differentiated service

Finally, we separately build F-ARIMA(3,0.4,5) model and S-ARIMA $(1,1,2)\times(0,1,1)_{10}$ model using the same data without differentiated service. The result is compared with MACTP model, as shown in Fig.4. S-ARIMA is slightly better than F-ARIMA, due to considering the periodicity of data. And it can be seen that distinguishing application types reduces error in some degree.

V. Conclusions

In order to solve the problem of integrated traffic prediction in the electric power data network that supports multi applications, a multi-applications comprehensive traffic prediction model is proposed in this paper. The simulation shows that MACPTP model is different from other predicting models. According to the flow characteristics of different application types, it considers factors, such as the trend, sudden and cyclical, reduces the prediction error in a certain extent, and improves the accuracy of prediction. After getting accurate traffic prediction, how to allocate bandwidth and other network resources reasonably and avoid network congestion will be the focus of the next research.

Acknowledgment

The paper is supported by the State Grid technical project “Research and Application of Service Quality Intelligent Probe and Holographic Estimation Orienting to Electric Power Data Network” (52010116000W).

References

[1] Yan Li, Li Zhao, Haibo Lin, A. C. Chow and J. R. Diamond, "A performance model for Fast Fourier Transform," Parallel & Distributed Processing, 2009. IPDPS 2009. IEEE International Symposium on, Rome, 2009, pp. 1-11.

[2] Y. h. Guo, X. p. Shi and X. d. Zhang, "A study of short term forecasting of the railway freight volume in China using ARIMA and Holt-Winters models," *Supply Chain Management and Information Systems (SCMIS), 2010 8th International Conference on*, Hong Kong, 2010, pp. 1-6.

[3] Babu,C.N.and B.E.Reddy.Predictive data mining on Average Global Temperature using variants of ARIMA models.in Advances in Engineering, Science and Management (ICAESM),2012 International Conference on. 2012.

[4] A. Salamanis, P. Meladianos, D. Kehagias and D. Tzovaras, "Evaluating the Effect of Time Series Segmentation on STARIMA-Based Traffic Prediction Model," *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, Las Palmas, 2015, pp. 2225-2230.

[5] YantaiShu,MinfangYu,JiakunLiuandO.W.W.Yang,"Wirelesstraffic modeling and prediction using seasonal ARIMA models," *Communications, 2003. ICC '03. IEEE International Conference on*, 2003, pp. 1675-1679 vol.3.

[6] C. Wang, Q. Wang, Y. Liu, Wang Su, Tian Lihui and Luo Ludong, "Applications of GIS to Power Distribution Dispatching and analysis of technical questions," *CICED 2010 Proceedings*, Nanjing, 2010, pp. 1-5.

[7] L. Li, X. Su, Y. Zhang, Y. Lin and Z. Li, "Trend Modeling for Traffic Time Series Analysis: An Integrated Study," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 6, pp. 3430-3439, Dec. 2015.

[8] R. Doorley, V. Pakrashi, B. Caulfield and B. Ghosh, "Short-term forecasting of bicycle traffic using structural time series models," *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, Qingdao, 2014, pp. 1764-1769.

[9] S. R. Yoo, J. S. Jeong, J. Y. Jong and G. K. Park, "Forecast of marine traffic volume using time series model," *2013 International Conference on Fuzzy Theory and Its Applications (iFUZZY)*, Taipei, 2013, pp