SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

April 2020, Vol. 63 142301:1–142301:8
https://doi.org/10.1007/s11432-019-2695-6

Prophet model and Gaussian process regression based user traffic prediction in wireless networks

Yu LI[1,2], Ziang MA[1,2], Zhiwen PAN[1,2*], Nan LIU[1,2] & Xiaohu YOU[1,2]

[1]*National Mobile Communications Research Laboratory, Southeast University, Nanjing* 210096*, China;*
[2]*Purple Mountain Laboratories, Nanjing* 211100*, China*

Received 16 June 2019/Revised 6 August 2019/Accepted 25 October 2019/Published online 9 March 2020

Abstract User traffic prediction is an important topic for wireless network operators. A user traffic prediction method based on Prophet and Gaussian process regression is proposed in this paper. The proposed method first employs discrete wavelet transform to decompose the user traffic time series to high-frequency component and low-frequency component. The low-frequency component bears the long-range dependence of user network traffic, while the high-frequency component reveals the gusty and irregular fluctuations of user network traffic. Then Prophet model and Gaussian process regression are applied to predict the two components respectively based on the characteristics of the two components. Experimental results demonstrate that the proposed model outperforms the existing time series prediction method.

Keywords wireless networks, traffic prediction, prophet model, Gaussian process regression

Citation Li Y, Ma Z A, Pan Z W, et al. Prophet model and Gaussian process regression based user traffic prediction in wireless networks. Sci China Inf Sci, 2020, 63(4): 142301, https://doi.org/10.1007/s11432-019-2695-6

1 Introduction

With the rapid development of mobile communication networks, the number of mobile devices surges, which leads to exponential growth of data traffic. Different mobile applications installed on mobile devices provide a wider variety of services for users. Meanwhile, users tend to consume more data traffic on diverse applications. The accurate portrayal of the user's network traffic behavior can not only help to improve the customized service for users, but also provide useful information for the network optimization, and hence it has attracted much research interests.

The analysis and understanding of mobile data traffic can be realized at aggregate level and at per-user level [1]. The mobile traffic prediction schemes at aggregate level are based on time series prediction models, for instance, seasonal auto regression integrated moving average (ARIMA) in [2], and machine learning algorithms, like neural networks (NN), Gaussian process (GP) in [3, 4]. Recently, more sophisticated methods like deep learning based approaches are employed to predict network traffic in [5]. The study of aggregate traffic can provide useful information for self-organization of the cellular network, for example, congestion control and energy saving based on the prediction results of the base station traffic [4, 6, 7].

Meanwhile, the research on per-user traffic prediction is relatively scarce. Existing studies mainly focus on descriptive statistics [8]. The per-user traffic in WLANs is analyzed and predicted using least minimum mean square estimation (LMMSE) in [9]. For per-user traffic in wireless mesh backbone networks, a prediction method based on deep belief network is proposed in [10], and a compressive sensing-based

* Corresponding author (email: pzw@seu.edu.cn)

© Science China Press and Springer-Verlag GmbH Germany, part of Springer Nature 2020 info.scichina.com link.springer.com

approach is proposed in [11]. The end-to-end network traffic for software defined networks (SDN) is estimated with a weighted-geometric-average process for fractal interpolation and cubic spline interpolation results of reconstructing the finer-granularity network traffic in [12]. Recently, per-user traffic in cellular network is analyzed and a Wavelet-ARMA based model is presented in [13]. According to [13], the wavelet transform is first applied to decompose the per-user traffic into two components that then are predicted via auto regression moving average (ARMA) model. The per-user traffic forecasting gives a finer understanding of network traffic which can help ISP (Internet service provider) to design suitable marketing plans based on the prediction results and to provide alarm mechanism to notify customers of data usage warning [14].

However, the main challenge of per-user traffic prediction is that the data usage of individual user is more stochastic compared with aggregated network traffic owing to emergency occurrences in people's ordinary life. Thus, user traffic is always bursty. To make matters worse, individual user tends to have large divergence in their behaviors. Some may not consume any traffic during some periods, while some use probably data all day long. Hence, it is difficult to capture the dynamic characteristics of per-user traffic and to predict traffic demand for individuals.

Motivated by this issue, a user network traffic prediction model based on Prophet model and Gaussian process regression (GPR) is proposed. The Prophet model is proposed by Facebook [15]. It is used to predict holiday traffic [16] and temporal performance measurements [17] in mobile cellular networks. GPR is a powerful machine learning method [17], which is used to predict spatial performance measurements [17] and network traffic in [3, 4]. Owing to more serious fluctuation in user's network traffic, existing time-series prediction algorithms cannot be directly applied to the network traffic prediction of individual user. Therefore, for the proposed method, wavelet transform is first employed to decompose the original user traffic data time series into low-frequency component and high-frequency component. The low-frequency component bears the long-range dependence of user network traffic, while the high-frequency component reveals the gusty and irregular fluctuations of user network traffic. Then Prophet model and GPR algorithm are used to predict the low-frequency component and high-frequency component respectively based on the characteristics of the two components.

The main contribution of this paper is that we propose a traffic prediction method based on Prophet model and Gaussian process regression which largely reduces the prediction error. Using the mean absolute percentage error as the evaluation metric, compared with the ARIMA model in [2] and Wavelet-ARMA model in [13], the proposed scheme can lower the prediction error to one third.

The remainder of the paper is organized as follows: Section 2 describes details of the dataset used and data preprocessing. The proposed prediction methodology is presented thoroughly in Section 3. Then Section 4 evaluates and discusses the experimental results. Three demand prediction models based on ARIMA, Wavelet-ARMA and the proposed method are examined. Experimental results are compared to demonstrate the superior performance of the proposed model. Finally, concluding remarks are made in Section 5.

2 Dataset

2.1 Dataset discription

The dataset used is an anonymized cellular data collected by an ISP between Sep. 1st to Oct. 31th. Each record consists of an anonymized IMEI (international mobile equipment identity), date, App name, start time of app usage, duration and the amount of data used in each usage. The dataset traces app usage of more than 70 thousand users in two months and the traffic consumption generated by each app usage.

2.2 Dataset preprocessing

Define the per-user traffic as the volume of data traffic a user used during a time period Δt. Hence, a per-user traffic time series of user can be denoted by

$$\boldsymbol{x}^i = [x_1^i, x_2^i, \ldots, x_t^i], \tag{1}$$

where $x_t^i = \log(1 + c_t^i)$ denotes the per-user demand within time slot $[t, t + \Delta t]$, c_t^i is the data traffic generated by user i at given time slot. Here, the base 10 logarithm operation is used to eliminate the divergence in magnitude. In this paper, we mainly study the traffic forecasting at a time scale of 1 h, namely, $\Delta t = 60$ min. With the definition of per-user traffic, the traffic forecasting is aimed to predict x_{t+1}^i based on its historical sequences $\boldsymbol{x}^i$.

3 Methodology

A per-user traffic prediction method based on Prophet model and Gaussian process is proposed. Discrete wavelet transform (DWT) is first employed to decompose the per-user traffic time series into high-frequency and low-frequency components. Then Prophet model and Gaussian process regression are applied to predict the two components respectively. In the end, the final prediction results can be obtained via inverse discrete wavelet transform (IDWT).

3.1 Data decomposition based on discrete wavelet transform

For user i, one-level wavelet decomposition is applied to per-user traffic time series $x_i(t) = \boldsymbol{x}^i$, obtaining scaling coefficients $c_i(n)$ and wavelet coefficients $d_i(n)$ as follows:

$$c_i(n) = \frac{1}{\sqrt{2}} \sum_{t=1}^{N} x_i(t) \varphi\left(\frac{t}{2} - n\right), \tag{2}$$

$$d_i(n) = \frac{1}{\sqrt{2}} \sum_{t=1}^{N} x_i(t) \psi\left(\frac{t}{2} - n\right), \tag{3}$$

where $\varphi(n)$ is scaling function and $\psi(n)$ is wavelet function. Assuming the length of time series $x_i(t)$ is N time slots, the length of the two components is $\frac{N}{2}$. Using inverse wavelet transformation, per-user demand time series $x_i(t)$ can be expanded by scaling function and wavelet function as follows:

$$x_i(t) = \sum_{n=1}^{N/2} c_i(n) \varphi\left(\frac{t}{2} - n\right) + \sum_{n=1}^{N/2} d_i(n) \psi\left(\frac{t}{2} - n\right). \tag{4}$$

The scaling coefficients $c_i(n)$ are the low-frequency component which exhibits the long-range dependence of the user traffic, while the wavelet coefficients $d_i(n)$ are high-frequency component which reveals the gusty and irregular fluctuations of user traffic. The predictions of two components $c_i(n)$ and $d_i(n)$ will be discussed in detail.

3.2 Low-frequency component prediction based on Prophet model

Prophet model is exploited to model and predict low-frequency component $c_i(n)$ because the low-frequency component has strong periodicity. In effect, the Prophet model tackles the forecasting problem as a curve-fitting issue. The idea of Prophet model is to decompose a time series (low-frequency component of the dataset in this paper) into three main components: trend, seasonality, and holiday as follows:

$$c_i(n) = g(n) + s(n) + h(n) + \varepsilon_n, \tag{5}$$

where $g(n)$ represents the non-periodic trend component, $s(n)$ represents periodic changes owing to yearly, weekly or daily periodicity, and $h(n)$ represents the impact of holidays or special events, ε_n is the error term assumed normally distributed that expresses the remaining changes which are not represented by the three main components. The task of Prophet model is to determine the three terms based on low-frequency component of the dataset.

A piecewise logistic growth model is adopted for modeling trend component:

$$g(n) = \frac{B(n)}{1+\exp(-(k+\boldsymbol{a}(n)^{\mathrm{T}}\boldsymbol{\delta})(n-(m+\boldsymbol{a}(n)^{\mathrm{T}}\boldsymbol{\gamma})))}, \tag{6}$$

where $B(n)$ is the expected capacities of the system, for example, the market sizes, k is the basic growth rate, $\boldsymbol{\delta}$ is the vector of rate adjustments at the change points, $\boldsymbol{a}(n) \in \{0,1\}^S$ is the vector of indicator function, m is the offset parameter, and $\boldsymbol{\gamma}$ is the parameter vector to make the function continuous. These parameters are time-varying. Suppose there are S change points happening at time slots $s_j, j = 1, \ldots, S$, then $a_j(n) = 1$ if $n \geqslant s_j$ otherwise $a_j(n) = 0$. $B(n)$ is a tuning parameter determined by the dataset. For simplicity and without loss of generality, here we choose the historical maximum value of per-user traffic as the value of $B(n)$. Fitting trend component requires estimating the parameters $k, \boldsymbol{\delta}$ and m. Assuming that k, m are a prior of Gaussian distribution, $\boldsymbol{\delta}$ is a prior of Laplace distribution, the parameters can be estimated by using maximum a posterior (MAP) estimation based on low-frequency component of the dataset.

Prophet model relies on Fourier series to model the periodic effects,

$$s(n) = \boldsymbol{e}(n)\boldsymbol{\beta} = \sum_{l=1}^{L}\left(a_l\cos\left(\frac{2\pi nl}{P}\right) + b_l\sin\left(\frac{2\pi nl}{P}\right)\right), \tag{7}$$

where

$$\boldsymbol{e}(n) = \left[\cos\left(\frac{2\pi(1)n}{P}\right), \ldots, \sin\left(\frac{2\pi(L)n}{P}\right)\right]. \tag{8}$$

L is the order of Fourier series. For weekly periodicity, the values of P is suggested as 7, and the value of L as 3. Fitting seasonality requires estimating the parameters $\boldsymbol{\beta} = [a_1, b_1, \ldots, a_L, b_L]^{\mathrm{T}}$. Assume $\boldsymbol{\beta}$ is a prior of Gaussian distribution, the parameter $\boldsymbol{\beta}$ can be estimated using a MAP estimator based on low-frequency component of the dataset.

Suppose there are M holidays, the holiday component is modeled as

$$h(n) = \boldsymbol{z}(n)\boldsymbol{\kappa} = \sum_{i=1}^{M}\kappa_i \cdot \mathbf{1}_{\{n\in D_i\}}, \tag{9}$$

where $\mathbf{1}_{\{n\in D_i\}} = 1$ if time slot n belongs to holiday i, otherwise $\mathbf{1}_{\{n\in D_i\}} = 0$, $\boldsymbol{z}(n) = [\mathbf{1}_{\{n\in D_1\}}, \ldots, \mathbf{1}_{\{n\in D_M\}}]$. And κ_i of a prior Gaussian distribution represents the influence of holiday i which can be estimated via MAP estimator based on low-frequency component of the dataset.

For convenient, combine seasonality and holiday effects for each observation into a matrix $\boldsymbol{X}_c = \{[\boldsymbol{e}(n) \quad \boldsymbol{z}(n)]\}_{n=1}^{N/2}$ and the change point indicator $\boldsymbol{a}(n)$ into a matrix $\boldsymbol{A} = \{\boldsymbol{a}(n)\}_{n=1}^{N/2}$. Define $\boldsymbol{\lambda} = (k, m, \boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{\kappa})$ as the parameters to be estimated for Prophet model. Then given the priors of all the parameters, the parameters estimation of the Prophet model is realized by a MAP estimator as follows:

$$\boldsymbol{\lambda}^{\mathrm{MAP}} = \operatorname{argmin}(-\log p(\boldsymbol{c}_i|\boldsymbol{X}_c, \boldsymbol{\lambda}) - \log p(\boldsymbol{\lambda})), \tag{10}$$

where $\boldsymbol{c}_i = \{c_i(n)\}_{n=1}^{N/2}$ is the low-frequency component of the per-user traffic, and $p(\boldsymbol{c}_i|\boldsymbol{X}_c, \boldsymbol{\lambda}) = N(\boldsymbol{\mu}_{ci}, \varepsilon)$ with

$$\boldsymbol{\mu}_{ci} = \frac{B}{(1+\exp(-(k+\boldsymbol{A}\boldsymbol{\delta})\cdot(n-(m+\boldsymbol{A}\boldsymbol{\gamma}))))} + \boldsymbol{X}_c\begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{\kappa}\end{bmatrix}. \tag{11}$$

Once the parameters $\boldsymbol{\lambda}^{\mathrm{MAP}}$ are estimated, the Prophet model are determined, the low-frequency component can be predicted as follows:

$$\hat{c}_i(n+1) = g(n+1) + s(n+1) + h(n+1). \tag{12}$$

3.3 High-frequency component prediction based on GPR

GPR is applied to model high-frequency components $d_i(n)$ which is quite stochastic. GPR tackles the time series prediction problem as a regression problem.

From high-frequency component, we can construct a training set with $N/2$ data points $\boldsymbol{X}_d = \{\boldsymbol{x}_n\}_{n=1}^{N/2}$ and corresponding targets $\boldsymbol{y} = \{d_i(1), d_i(2), \ldots, d_i(N/2)\}^{\mathrm{T}}$, where $\boldsymbol{x}_n = [d_i(n-3), d_i(n-2), d_i(n-1)]$ is input feature vector and $y_n = d_i(n)$ is the corresponding observation which is the high-frequency component of the dataset in this paper. In GPR, it is assumed that each observation is generated from a latent function f and addictive Gaussian noise ω, namely,

$$y_i = f(\boldsymbol{x}_i) + \omega, \tag{13}$$

where $\omega \sim N(0, \sigma_n^2)$ is zero mean Gaussian noise with variance σ_n^2. And a prior distribution of the latent function f is defined as

$$p(\boldsymbol{f}|\boldsymbol{X}_d) = N(\boldsymbol{0}, \boldsymbol{K}), \tag{14}$$

which denotes the multivariate normal distribution with zero mean, covariance matrix $\boldsymbol{K}$ (also named as kernel). The covariance matrix $\boldsymbol{K}$ is obtained from a kernel function k, namely $\boldsymbol{K}_{ij} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$. Here the squared exponential co-variance function

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\frac{1}{2\theta^2}(\boldsymbol{x}_i - \boldsymbol{x}_j)^{\mathrm{T}}(\boldsymbol{x}_i - \boldsymbol{x}_j)\right) \tag{15}$$

is chosen as kernel function. θ is kernel hyper-parameters that can be estimated by maximizing the marginal likelihood $p(\boldsymbol{y}|\boldsymbol{X}_d, \theta) = N(\boldsymbol{0}, \boldsymbol{K}_T)$, where $\boldsymbol{K}_T = \boldsymbol{K} + \sigma_n^2\mathbf{I}$ represents a noisy covariance matrix. The maximum likelihood estimation of θ is

$$\theta^{ML} = \operatorname{argmin}(-\log p(\boldsymbol{y}|\boldsymbol{X}_d, \theta)). \tag{16}$$

Once the hyper-parameter θ is estimated, the predicted value of test data $\boldsymbol{x}_* = [d_i(n-2), d_i(n-1), d_i(n)]$ can be calculated from the predictive distribution $p(y_*|\boldsymbol{x}_*, \theta)$ which is Gaussian with mean μ_* and variance σ_*^2, and

$$\mu_* = \boldsymbol{k}_*\boldsymbol{K}_T^{-1}\boldsymbol{y}, \tag{17}$$

where $\boldsymbol{k}_* = (k(\boldsymbol{x}_*, \boldsymbol{x}_1), \ldots, k(\boldsymbol{x}_*, \boldsymbol{x}_{N/2}))^{\mathrm{T}}$. The mean μ_* is usually used as prediction for test data $\boldsymbol{x}_*$. Namely,

$$\hat{d}_i(n+1) = \mu_*. \tag{18}$$

3.4 Per-user traffic prediction

Once the high-frequency component and the low-frequency component are predicted by Prophet model and GPR, inverse discrete wavelet transform is applied to obtain the final prediction result of per-user traffic as

$$\hat{x}_i(t+1) = \sum_{n=1}^{N/2} \hat{c}_i(n+1)\varphi\left(\frac{t+1}{2} - n - 1\right) + \sum_{n=1}^{N/2} \hat{d}_i(n+1)\psi\left(\frac{t+1}{2} - n - 1\right). \tag{19}$$

The above algorithm can be summarized in Algorithm 1.

4 Experiments

4.1 Performance metrics

To evaluate the performance of the models, the root mean square error (RMSE) and the mean absolute percentage error (MAPE) are used. A smaller value of MAPE and RMSE indicates a more effective prediction model.

Algorithm 1 Traffic prediction algorithm

Input: Per-user traffic time series $x_i(t)$.
Output: Prediction $\hat{x}_i(t+1)$.
for User $i = 1$ to P **do**
  $c_i(n) = \frac{1}{\sqrt{2}}\sum_{t=1}^{N} x_i(t)\varphi(\frac{t}{2}-n), d_i(n) = \frac{1}{\sqrt{2}}\sum_{t=1}^{N} x_i(t)\psi(\frac{t}{2}-n)$;
  for time slot $n = 1$ to $N/2$ **do**
    $g(n) = \frac{B(n)}{1+\exp(-(k+\boldsymbol{a}(n)^{\mathrm{T}}\boldsymbol{\delta})(n-(m+\boldsymbol{a}(n)^{\mathrm{T}}\gamma)))}$;
    $s(n) = \boldsymbol{e}(n)\boldsymbol{\beta} = \sum_{l=1}^{L}(a_l\cos(\frac{2\pi nl}{P}) + b_l\sin(\frac{2\pi nl}{P}))$;
    $h(n) = \boldsymbol{z}(n)\boldsymbol{\kappa} = \sum_{i=1}^{M}\kappa_i\cdot\mathbf{1}_{\{n\in D_i\}}$;
  end for
  $\boldsymbol{X}_c = \{[\boldsymbol{e}(n)\ \ \boldsymbol{z}(n)]\}_{n=1}^{N/2}, \boldsymbol{c}_i = \{c_i(n)\}_{n=1}^{N/2}, \boldsymbol{A} = \{\boldsymbol{a}(n)\}_{n=1}^{N/2}$;
  $\boldsymbol{\lambda} = (k, m, \boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{\kappa}), p(\boldsymbol{c}_i|\boldsymbol{X}_c, \boldsymbol{\lambda}) = N(\boldsymbol{\mu}_{ci}, \varepsilon), \boldsymbol{\mu}_{ci} = \frac{B}{(1+\exp(-(k+\boldsymbol{A\delta})\cdot(n-(m+\boldsymbol{A\gamma}))))} + \boldsymbol{X}_c[\begin{smallmatrix}\boldsymbol{\beta}\\ \boldsymbol{\kappa}\end{smallmatrix}]$;
  $\boldsymbol{\lambda}^{\mathrm{MAP}} = \mathrm{argmin}(-\log\ p(\boldsymbol{c}_i|\boldsymbol{X}_c, \boldsymbol{\lambda}) - \log\ p(\boldsymbol{\lambda}))$;
  $\hat{c}_i(n+1) = g(n+1) + s(n+1) + h(n+1)$;
  for time slot $n = 1$ to $N/2$ **do**
    $\boldsymbol{x}_n = [d_i(n-3), d_i(n-2), d_i(n-1)]$;
    $y_n = d_i(n)$;
  end for
  $\boldsymbol{X}_d = \{\boldsymbol{x}_n\}_{n=1}^{N/2}, \boldsymbol{y} = \{y_n\}_{n=1}^{N/2}$;
  $p(\boldsymbol{y}|\boldsymbol{X}_d, \theta) = N(\mathbf{0}, \mathbf{K}_T), \boldsymbol{K}_T = \boldsymbol{K} + \sigma_n^2\mathbf{I}, k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp(-\frac{1}{2\theta^2}(\boldsymbol{x}_i - \boldsymbol{x}_j)^{\mathrm{T}}(\boldsymbol{x}_i - \boldsymbol{x}_j))$;
  $\theta^{ML} = \mathrm{argmin}(-\log\ p(\boldsymbol{y}|\boldsymbol{X}_d, \theta))$;
  $\boldsymbol{x}_* = [d_i(n-2), d_i(n-1), d_i(n)], \boldsymbol{k}_* = (k(\boldsymbol{x}_*, \boldsymbol{x}_1), \ldots, k(\boldsymbol{x}_*, \boldsymbol{x}_{N/2}))^{\mathrm{T}}$;
  $\hat{d}_i(n+1) = \boldsymbol{k}_*\boldsymbol{K}_T^{-1}\boldsymbol{y}$;
  $\hat{x}_i(t+1) = \sum_{n=1}^{\frac{N}{2}}\hat{c}_i(n+1)\varphi(\frac{t+1}{2}-n-1) + \sum_{n=1}^{\frac{N}{2}}\hat{d}_i(n+1)\psi(\frac{t+1}{2}-n-1)$.
end for

Definitions of these metrics are

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(|y_i - \hat{y}_i|)^2}, \tag{20}$$

$$\mathrm{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i}. \tag{21}$$

4.2 Experimental results

In this paper, we compare the proposed model with traditional time series prediction method ARIMA and wavelet-ARMA proposed in [13]. The length of training set is 7 days from first September to 7th September, and the aim is predicting the data usage of the next 24 h. Figure 1 shows the predicted traffic of a user.

Forecast resolution and forecast horizon are two important parameters for a forecasting problem. Forecast resolution relies on the time scale that the data are aggregated. Figures 2 and 3 illustrate the performance of prediction methods for different forecast resolution. It is obvious that predicting user traffic in a short time period is more difficult because of strong randomness of human behavior. Therefore, the larger the forecast resolution, the better the performance. For all different forecast resolutions, the proposed model outperforms the ARIMA model and wavelet-ARIMA model. Even at time scale of 15 min, the proposed model can achieve a MAPE of less than 60%, compared with ARIMA model and wavelet-ARIMA model, the values of MAPE are nearly 100%.

Forecast horizon means the time range that the forecasting method predicts. Figure 4 presents the performance of three methods for different forecast horizons. It shows that the proposed method has the best performance for predicting the next 24 h traffic. For 24 h prediction, the proposed model achieves a RMSE of 1.18, compared with 3.5 and 3.74 of other two models. With the increase of the forecast horizon, the prediction error becomes larger and finally tends to be stable.

1000 users are chosen randomly from the dataset, and user traffic for the next 24 h is predicted. Figure 5 shows the cumulative distribution function (CDF) of RMSE of ARIMA, wavelet-ARIMA model

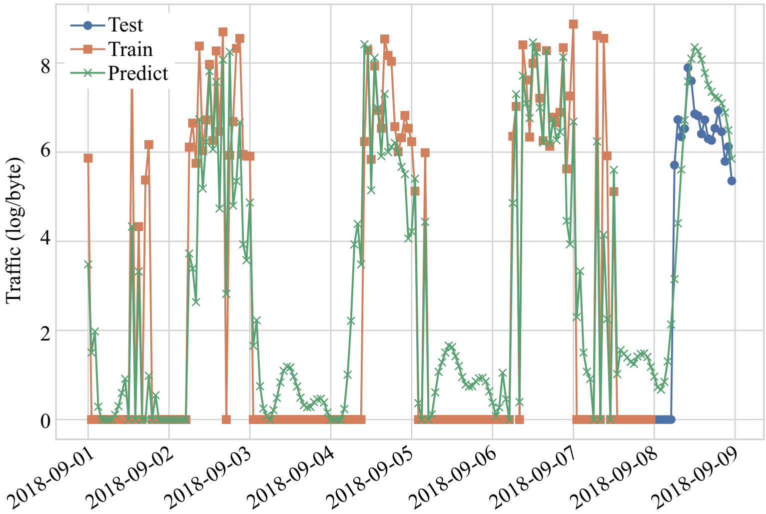

Figure 1 (Color online) Prediction result for a user.

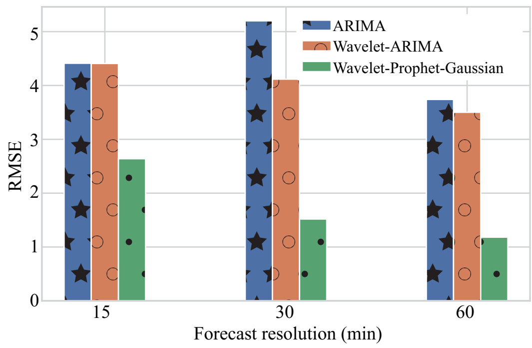

Figure 2 (Color online) RMSE of three methods for different forecast resolutions.

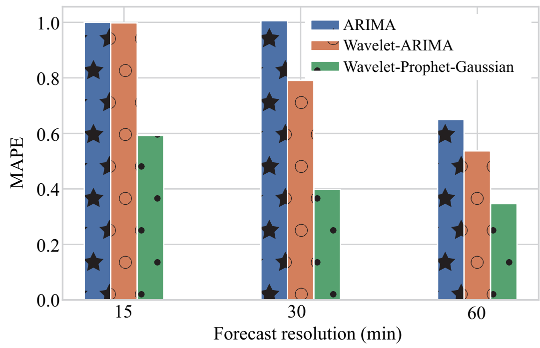

Figure 3 (Color online) MAPE of three methods for different forecast resolutions.

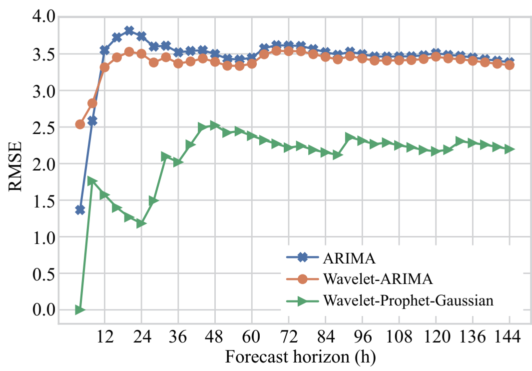

Figure 4 (Color online) RMSE of three methods for different forecast horizons.

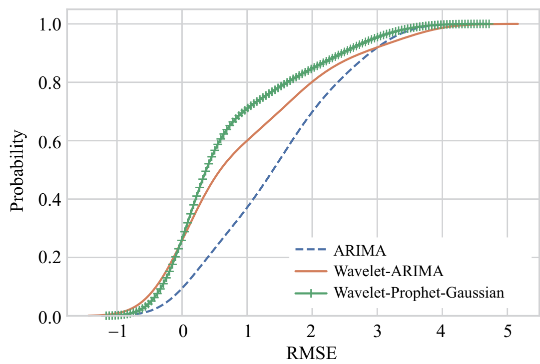

Figure 5 (Color online) CDF of RMSE for 1000 users.

and the proposed model. Figure 5 shows that the predictions results of nearly 80% users have an RMSE lower than 1 based on the proposed model, while only 60% and 40% users have a RMSE lower than 1 for wavelet-ARIMA model and ARIMA model, respectively.

Considering the computing overheads of the proposed model, the Gaussian process model has a time complexity of $O(n^3)$ and space complexity of $O(n^2)$, where n is the number of data samples. Compared with ARIMA model which has a time complexity of $O(n)$, the proposed model has more computing overheads. However, the value of n is usually small. Supposing we use a week's historical data to forecast user hourly traffic, the value of n is 24×7. For example, for training and predicting one hundred users, the proposed model takes 310 s. While ARIMA model takes 246 s. However, the proposed algorithm can largely lower the prediction error.

5 Conclusion

In this paper, we proposed a user traffic prediction method based on Prophet model and Gaussian process regression model. Discrete wavelet transform was employed to decompose the traffic time series to high-frequency component and low-frequency component, then Prophet model and GPR were used to predict the two components respectively. Compared with ARIMA and wavelet-ARIMA model, the proposed model can reduce RMSE to one third.

Acknowledgements This work was partially supported by National Key Research and Development Project (Grant No. 2018YFB1802402) and Huawei Tech. Co., Ltd.

References

1 Naboulsi D, Fiore M, Ribot S, et al. Large-scale mobile traffic analysis: a survey. IEEE Commun Surv Tut, 2016, 18: 124–161
2 Xu F L, Lin Y Y, Huang J X, et al. Big data driven mobile traffic understanding and forecasting: a time series approach. IEEE Trans Serv Comput, 2016, 9: 796–805
3 Le L, Sinh D, Tung L, et al. A practical model for traffic forecasting based on big data, machine-learning, and network KPIs. In: Proceedings of IEEE Annual Consumer Communications and Networking Conference (CCNC), Las Vegas, 2018. 1–4
4 Le L, Sinh D, Lin B P, et al. Applying big data, machine learning, and SDN/NFV to 5G traffic clustering, forecasting, and management. In: Proceedings of IEEE Conference on Network Softwarization and Workshops (NetSoft), Montreal, 2018. 168–176
5 Fang L Y, Cheng X, Wang H N, et al. Mobile demand forecasting via deep graph-sequence spatiotemporal modeling in cellular networks. IEEE Internet Things J, 2018, 5: 3091–3101
6 Dawoud S, Uzun A, Göndör S, et al. Optimizing the power consumption of mobile networks based on traffic prediction. In: Proceedings of IEEE 38th Annual Computer Software and Applications Conference, Vasteras, 2014. 279–288
7 Hu J M, Heng W, Zhang G D, et al. Base station sleeping mechanism based on traffic prediction in heterogeneous networks. In: Proceedings of International Telecommunication Networks and Applications Conference (ITNAC), Sydney, 2015. 83–87
8 Yang J, Qiao Y Y, Zhang X Y, et al. Characterizing user behavior in mobile Internet. IEEE Trans Emerg Top Comput, 2015, 3: 95–106
9 He G H, Hou J C, Chen W-P, et al. Characterizing individual user behaviors in wlans. In: Proceedings of 10th ACM Symposium on Modeling, Analysis, and Simulation of Wireless and Mobile Systems (MSWiM), Chania, 2007. 132–137
10 Nie L, Jiang D D, Yu S, et al. Network traffic prediction based on deep belief network in wireless mesh backbone networks. In: Proceedings of IEEE Wireless Communications and Networking Conference (WCNC), San Francisco, 2017. 1–5
11 Jiang D D, Wang W J, Shi L, et al. A compressive sensing-based approach to end-to-end network traffic reconstruction. IEEE Trans Netw Sci Eng, 2020, 7: 507–519
12 Huo L W, Jiang D D, Zhu X N, et al. A SDN-based fine-grained measurement and modeling approach to vehicular communication network traffic. Int J Commun Syst, 2019, 56: e4092
13 Wu J, Zeng M, Chen X L. Characterizing and predicting individual traffic usage of mobile application in cellular network. In: Proceedings of ACM International Joint Conference and International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers (UbiComp18), New York, 2018. 852–861
14 Lai Y-T, Wu Y-P, Yu C-H, et al. Mobile data usage prediction system and method. In: Proceedings of 31st International Conference on Advanced Information Networking and Applications Workshops (WAINA), Taipei, 2017. 484–486
15 Taylor S J, Letham B. Forecasting at scale. Am Stat, 2018, 72: 37–45
16 Xu M, Wang Q L, Lin Q L. Hybrid holiday traffic predictions in cellular networks. In: Proceedings of IEEE/IFIP Network Operations and Management Symposium, Taipei, 2018. 1–6
17 Riihijarvi J, Mahonen P. Machine learning for performance prediction in mobile cellular networks. IEEE Comput Intell Mag, 2018, 13: 51–60