Prediction of Traffic Density and Interest Using Real Time Mobile Traffic Data

Yuan Gao[1,2,4], Ao Hong[1], Quan Zhou[1], Xiangyang Li[4],Shasha Liu[3,5,*] and Bingjia Shao[3]
[1] Xichang Satellite Launch Center, Xichang, 615000, China
[2] State Key Laboratory on Microwave and Digital Communications,
National Laboratory for Information Science and Technology, Tsinghua University, Beijing, 100084, China
[3] School of Economics and Business Administration, Chongqing University, Chongqing, China
[4] China Defense Science and Technology Information Center, Beijing, 100030, China
[5] Chongqing Jadun Economics and Management Academy, Chongqing, China
*Corresponding Author: liushasha@cqu.edu.cn

***Abstract*—With the increasing number of mobile users and applications, data services and on-demand suggestions gradually play an important role in mobile life. In this work, based on the wireless personal communication network, we propose a novel prediction method of user traffic density and personal interest based on real time mobile traffic data and wireless positioning information. Service provider could send precise information to target users based on the prediction to ensure the quality and push service. Then we execute a simulation using real traffic data around Tsinghua University obtained from network service provider, and evaluate the accuracy of the prediction. Simulation results indicate that, our proposed method by using the traffic data and position jointly can significantly increase the rate of successful recommendation.**

Keywords—traffic density; user interest; big data prediction; real time; traffic data

I. INTRODUCTION

In recent years, the amount of mobile terminals is increasing at a tremendous speed, which lead to the burden of mobile base stations, especially when the multimedia terminals with many applications installed. Such terminals will suggest and push information to users according to the historical data and machine learning results, which will bring heavy load to mobile networks.

So network operators and application developers are eagle to reduce the unnecessary push of information, which is the key to reduce the load of the network. So both industry and academic researchers are trying to propose the precise information that the user will require. In [1], the author presents the study of business scholarship using big data, but there is no theoretical analysis to describe the exact model. So in [2] and [3], the authors give the prediction of user mobility according to the record of public WLAN or 4G wireless networks, the mobility prediction is useful to decide whether the user is interested in given area, but only prediction of mobility could not give exact information of push information, so in [4], the prediction based on big data analysis is given, which makes it more precise to perform the prediction, however, the method to obtain wireless big data is the key problem that limit the utilization of the analysis. In [5], the authors present the method to discover the friendship using social networks, based on social applications, users could be sorted and grouped. Then in [6] and [7], push information based on demand of users is proposed, users will not receive recommendations without permission, this is also the method to reduce signaling or traffic load. To make the information clear and precisely, researchers in references [8-10] show the researches about historical data analysis and prediction of upcoming events, but this will allow the interactions of users to share their personal data, which may cause the leak of privacy. So in [10-12], prediction of mobility based on public data has been proposed. In these researches, information that users perform establishment to base stations has been grasped, such as signaling and public broadcast information, so the trace of mobility could be detected. And in [13-15], the improved prediction of user mobility has been proposed to make it more clear and stable, both speed and direction are predicted.

In this paper, aimed at the need to send precise recommendation to target users, and reduce unnecessary information to avoid load to wireless resources, we propose a novel method to predict traffic density and user interest, which may be the basis to send recommendations and suggestions to target users. The real time traffic data and position information will be collected at the base station side, no other information is required according to this method.

The rest of the paper is organized as follows. In part Ⅱ, we give the system model of traffic density and interest in wireless networks. In part Ⅲ, we present our novel design using real time traffic data. In part Ⅳ, simulation about the proposed method is given using the real data from mobile wireless base stations. Conclusions are given in the last part.

Dr. Yuan Gao is currently an assistant professor with the Department of Electronic Engineering, Tsinghua University and China Defense Science and Technology Information Center. The corresponding author, Dr. Shasha Liu is with School of Economics and Business Administration, Chongqing University, who contribute the same as the first author.

978-1-5090-5952-2/16 $31.00 © 2016 IEEE
DOI 10.1109/IIKI.2016.31

II. SYSTEM MODEL OF WIRELESS INFORMATION PUSH

Recent wireless personal communication systems play an important role in our daily life. The mobile terminals are equipped with many sensors and functions, such as GPS, fingerprint sensors, NFC, etc. Varieties of sensors consist of hybrid information space, which could provide necessary elements for service providers and network operators to push advertised information.

In figure 1, we present the map of Tsinghua University with 3 base stations located in the campus. The red line indicates the border of the campus, and three base station (painted purple) will work in LTE-Advanced Release 11 mode (typically 4G, China Mobile). The three base stations will serve about 10 thousand persons at the most.

Fig.1. Map of Tsinghua University with 3 base stations located.

To make a successful necessary push, we must infer two things: where is the target user and what is the interest of the target user. To indicate the position of the user, we have many methods, such as GPS, network based TOA, DOA, TDOA, etc.

In this work, we consider the method using network assisted positioning method using route information. Network assisted positioning method is widely known as A-GPS, equipped in many mobile terminals, we do not introduce this in this work, detailed information could be found in [16]. Coupled with route information, that means we assume that all the users are moving through the defined road.

In the following work, we first need to define traffic density. Traffic density is used to define the load of the network, if the density is large, the base station will work in high load. In equation (1), N means the users located in the area of base stations and S is the coverage of base stations.

$$\rho = \frac{N}{S} \tag{1}$$

III. PREDICTION OF TRAFFIC DENSITY AND USER INTEREST

In this part, prediction of traffic density and user interest will be given in two different sub-sections.

A. Prediction of Traffic Density

Different types of traffic indicate the different requirements of users in mobile networks. If a method $F(t_k)$ is to forecast the future using the previous time observations, we call this the prediction operation, t_k is the k^{th} observation period. The prediction error is defined using the following equation:

$$e(T,k) = \frac{\| F(t_k) - V(t_k) \|}{V(t_k)} \tag{2}$$

Where $e(T,k)$ is the actual absolute value of real data $V(t_k)$ and predicted data $F(t_k)$. To predict the future traffic density using the previous status, we have the following method called exponential smoothing. Define the observation sequence $\{x_t\} = \{x_t(0), x_t(1), \cdots, x_t(\mathrm{L})\}$, which means the received traffic strength at base station side with observation period L. $\{b_t\}$ is the predicted sequence obtained using our proposed method and $\{c_t\}$ is the expected proportion of the forecasted density. For any given time t, we will get the prediction sequence $F(t+m)$ where m is the predicted length:

$$s_t = \alpha \frac{x_t}{c_{t-L}} + (1-\alpha)(s_{t-1} + b_{t-1}) \tag{3}$$

$$b_t = \beta\left[s_t - s_{t-1} + (\alpha - \beta)b_{t-1}\right] \tag{4}$$

$$c_t = \gamma \frac{x_t}{s_t} + (1-\gamma)c_{t-L} \tag{5}$$

In the above equations, $\alpha \in [0,1]$ is the factor to change the sensitivity of s_t, $\beta \in [0,1]$ and $\gamma \in [0,1]$ are the factors to change the sensitivity of b_t and c_t.

B. Prediction of User Interest

The traffic density indicates the types of interest that the users will require, and the interest of users of any given group will be predicted to decide whether the base station will push information for target user group.

By analyzing the type of traffic M, the interest for any given user i is defined as $\zeta_m(i)$, where this interest is the probability that the user need this information. To indicate the accuracy of the demand, we use the outage probability to define the prediction. The outage probability is defined that the probability that the base station pushes the target information to target user but the user does not need this information:

$$p_{blk}(t) = \sum_{i \in I} prob\{m(i)\} \times prob\{\zeta_m(i)\} \tag{6}$$

If the prediction is wrong, the user will reject the information and cause a failure. To indicate the effectiveness of our method, the outage probability is the key indicator. So we have the following method to predict the user interest among all type of traffic.

In the proposed method given in table 1, we first initialize all the parameters for traffic prediction using the exponential smoothing method, then for every given observation time t, we compute the outage probability for every types of traffic, and if the probability is lower than fixed value p_{static}, then we consider it a sign to push the information, else we will not push the information. Through this method, we could choose whether to send recommended information at the base station side to make an accurate recommendation.

Proposed Algorithm:
1. Initialize $I, M, F, \alpha, \beta, \gamma$
2. Initialize L, let $t = 0$.
3. **Repeat:**
4. $t = t + 1$;
5. Get $V(t_k)$ and $F(t_k)$ by solving equation (2)-(5) ;
6. Compute $p_{blk}(t)$
7. **If** $p_{blk}(t) \le p_{static}$, PUSH information
8. **Else Reject PUSH**
9. **End**

IV. Simulation and Analysis

In this part, we present a simulation to indicate the effectiveness of our proposed work [17] [18]. The data of wireless information is obtained through the three test base stations located in Tsinghua University.

We choose 5 types of traffic among all collected data, they are: online video, WeChat/QQ, Online Course, Academic Surfing and shopping, which takes up about 83% of all the traffic.

Fig. 2 The predicted push area with different 5 types of traffic. Different colors indicate different probability of interest.

In figure 2, we show the result of predicted interest area in the map of Tsinghua University. There are 5 types of traffic. From left to right and up to bottom, they represent WeChat/QQ, Online Video, Online Course, Academic Surfing and shopping. Different colors indicate different types of interest, deep red located as the center of the area means the interest is larger than 90%, light red means $p \in (0.7, 0.9]$, yellow means the probability means $p \in (0.5, 0.7]$, green means $p \in (0.3, 0.5]$, cyan means $p \in (0.1, 0.3]$ and blue means the interest is lower than 10%. The distribution is greatly related to the area and time, for example, online video is happened in the dormitory area, that means users located at such area are willing to watch online video, which is consistent with the properties of such area. In our proposed method, we will make a precise prediction of the behavior of users, to allow the push of the information. In figure 3, we evaluate the prediction error of our proposed method.

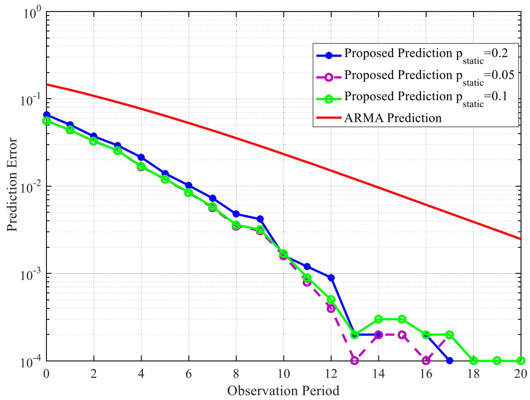

Fig. 3 Prediction Error of the proposed method compared to the scenario without prediction.

In figure 3, the x-axis means the observation period in second, and the y-axis is the prediction error. Different colors mean different method. The red line means the prediction using ARMA method, the others indicate our proposed methods, blue means $p_{static} = 0.2$, green means $p_{static} = 0.1$ and purple means $p_{static} = 0.05$. It is obvious that the reference ARMA method performs the worst. It is because this method does not consider the correlation of time and space jointly, one dimension is missed. The other three lines indicate our proposed methods. With the increasing of observation period, the prediction error becomes lower, this is because the length of observation period determines the effect of prediction, based on the assumption that the data traffic is stationary in a short time. p_{static} is the bound that decide the outage probability to push the information, from the simulation, we can infer that, when this value is lower than 0.1, the influence is not notable. When $p_{static} = 0.2$, it is a little bit worse than other two bounds, physically speaking, it reflects the difference of coverage of circle in figure 2, $p_{static} \le 0.1$ means the coverage defined by deep red and $p_{static} = 0.2$ is the circle of light red. So in engineering point of view, $p_{static} = 0.1$ is a suitable value to adopt the proposed method and reaches the balance of computing complexity and effect.

V. CONCLUSION

In this paper, we describe a novel method to predict traffic density and user interest jointly in wireless personal communication systems, which is effective and reliable to make accurate push to target users. Our proposed method could detect the major types of data traffic and make prediction using real time and historical traffic data, the probability that where the target area need accurate information push is simulated and the prediction error is also evaluated. Our proposed method is proved effective to predict the interest of users to make sure the information push is reliable.

ACKNOWLEDGMENT

This work is funded by China's 973 project under grant of 2012CB316002 and China's 863 project under grant of 2013AA013603, 2012AA011402, National Natural Science Foundation of China(61201192), The Open Research Fund of National Mobile Communications Research Laboratory, Southeast University (2012D02); International Science and Technology Cooperation Program(2012DFG12010); National S & T Major Project (2013ZX03001024-004) ,Operation Agreement Between Tsinghua University and Ericsson, Qualcomm Innovation Fellowship, whose funding support is gratefully acknowledgment. The author would also like to thank all the reviewers; their suggestions help improve the work a lot.

REFERENCES

[1] Frizzo-Barker, Julie, et al. "An empirical study of the rise of big data in business scholarship." International Journal of Information Management 36.3(2016):403-413..

[2] T. V. T. Duong and D. Q. Tran, "Mobility prediction based on collective movement behaviors in public WLANs," Science and Information Conference (SAI), 2015, London, 2015, pp. 1003-1010.

[3] W. Sun, D. Miao, X. Qin and G. Wei, "Characterizing User Mobility from the View of 4G Cellular Network," 2016 17th IEEE International Conference on Mobile Data Management (MDM), Porto, Portugal, 2016, pp. 34-39.

[4] Y. Qiao; Y. Cheng; J. Yang; J. Liu; N. Kato, "A Mobility Analytical Framework for Big Mobile Data in Densely Populated Area," in IEEE Transactions on Vehicular Technology , vol.PP, no.99, pp.1-1

[5] Jianwei Niu, Danning Wang and Jie Lu, "Mining friendships through spatial-temporal features in mobile social networks," 2015 IEEE 34th International Performance Computing and Communications Conference (IPCCC), Nanjing, 2015, pp. 1-8.

[6] Z. Yang, B. Wu, K. Zheng, X. Wang and L. Lei, "A Survey of Collaborative Filtering-Based Recommender Systems for Mobile Internet Applications," in IEEE Access, vol. 4, no. , pp. 3273-3287, 2016.

[7] Schneider, Matthew J., S. Gupta, and R. J. Hyndman. "Forecasting sales of new and existing products using consumer reviews: A random projections approach." Social Science Electronic Publishing 32.2(2015):243-256.

[8] J. Zhang, X. Pan, M. Li and P. S. Yu, "Bicycle-Sharing System Analysis and Trip Prediction," 2016 17th IEEE International Conference on Mobile Data Management (MDM), Porto, Portugal, 2016, pp. 174-179.

[9] P. Yu, "Recommendation method for mobile network based on user characteristics and user trust relationship," 2016 IEEE International Conference on Big Data Analysis (ICBDA), Hangzhou, China, 2016, pp. 1-6.

[10] R. Margolies et al., "Exploiting Mobility in Proportional Fair Cellular Scheduling: Measurements and Algorithms," in IEEE/ACM Transactions on Networking, vol. 24, no. 1, pp. 355-367, Feb. 2016.

[11] Z. Xing, H. Tian, T. Chen and J. Zhang, "Feature tendency based location prediction in LBSNs," 2016 18th International Conference on Advanced Communication Technology (ICACT), Pyeongchang, 2016, pp. 665-671.

[12] A. Nadembega, A. S. Hafid and R. Brisebois, "Mobility prediction model-based service migration procedure for follow me cloud to support QoS and QoE," 2016 IEEE International Conference on Communications (ICC), Kuala Lumpur, Malaysia, 2016, pp. 1-6.

[13] C. Qi, Z. Zhao, R. Li and H. Zhang, "Characterizing and Modeling Social Mobile Data Traffic in Cellular Networks," 2016 IEEE 83rd Vehicular Technology Conference (VTC Spring), Nanjing, 2016, pp. 1-5.

[14] F. Tang, M. Guo, S. Guo and C. Z. Xu, "Mobility Prediction Based Joint Stable Routing and Channel Assignment for Mobile Ad Hoc Cognitive Networks," in IEEE Transactions on Parallel and Distributed Systems, vol. 27, no. 3, pp. 789-802, March 2016.

[15] A. Li, Q. Lv, Y. Qiao and J. Yang, "Improving mobility prediction performance with state based prediction method when the user departs from routine," 2016 IEEE International Conference on Big Data Analysis (ICBDA), Hangzhou, China, 2016, pp. 1-7.

[16] A. Alkhateeb and D. Aloi, "Polarization performance study of a GPS antenna mounted on a small t-tail aircraft in landing position using EM modeling," Digital Information, Networking, and Wireless Communications (DINWC), 2015 Third International Conference on, Moscow, 2015, pp. 39-43.

[17] M. Dash, K. K. Koo, S. P. Krishnaswamy, Y. Jin and A. Shi-Nash, "Visualize People's Mobility - Both individually and Collectively - Using Mobile Phone Cellular Data," 2016 17th IEEE International Conference on Mobile Data Management (MDM), Porto, Portugal, 2016, pp. 341-344.

[18] B. Yang; W. Guo; B. Chen; G. Yang; J. Zhang, "Estimating Mobile Traffic Demand using Twitter," in IEEE Wireless Communications Letters , vol.PP, no.99, pp.1-1