



Intelligent User Profile Prediction in Radio Access Network

Yaxing Qiu^(✉), Xidong Wang^(✉), Fengjun Wang^(✉),
and Sen Bian^(✉)

Green Communication Research Center of China Mobile Research Institute,
CMCC, Beijing, People's Republic of China
{qiuyaxing, wangxidong, wangfengjun, biansen}
@chinamobile.com

Abstract. With increasingly dense deployment of base stations, power consumption, resource utilization, interference management, and user experience optimization in mobile network have become more and more important challenges. The effective prediction of temporal and spatial data traffic distribution, en-powered by intelligent user profile prediction, will be essential. This paper investigates user mobility and user service patterns prediction by generating refined user trajectory and 2-D service feature models. Specifically, an improved Bayesian trajectory prediction strategy coupled with time-domain features extraction used for user service pattern prediction is proposed. Large scale field test using 2300 active mobile devices (users) across 1600 cells in a live network showed promising results of 85% trajectory prediction accuracy and 70% service pattern prediction accuracy. The accurate prediction of user-level behavior pattern is of great significance not only for the improvement of network energy-efficiency, but also for the guarantee of user experience and the optimization of network utilization.

Keywords: Mobility prediction · Service pattern prediction · Machine learning

1 Introduction

As one of the biggest telecom operators in the world, China Mobile has the widest wireless access networks in China, in which the total number of base stations (BSs) has exceeded 3 million by 2016 [1]. For the 5th Generation communication system (5G), more BSs will continue to be installed to support up to 1,000-fold gains in capacity. With increasingly dense deployment of base stations, wireless networks have become more and more advanced and complicated. The wireless networks are generating a large amount of wireless data (such as measurement report, traffic load, resource usages, etc.) all the time, and this drives the wireless communication networks into the era of big data [2]. In practice, the traffic load and resource usages of network are closely related to the behavior pattern of the mobile users. Analyzing users' behavior pattern based on their interaction with wireless networks is a key problem in network usage mining. Also, user mobility and behavior patterns prediction can significantly

benefit the resource-constrained network automation, from network planning, network traffic monitoring to network management in mobile networks [3].

Mobility is an inherent characteristic of users in mobile networks. User mobility prediction allows estimating or predicting the location and trajectory of users at the next moment. Nowadays, a number of studies have been proposed to conduct human mobility pattern mining based on human mobility data in many types of wireless networks. In [4], Order-k Markov predictor is used to predict next place on the basis of the sequence of the k-most latest locations in users' trajectories. Additionally, some prediction methods using machine learning techniques have been suggested as useful approaches to improve the prediction accuracy. One advantage of the learning-based model is that mobile contexts can be quantitatively measured and mapped into a feature space for prediction. Tomar and Verma have suggested the trajectory prediction method using a Support Vector Machine (SVM) and used it as the regression analysis method in [5]. Furthermore, deep learning approach for spatiotemporal modeling and prediction in cellular networks is also proposed with using big system data in [6]. Unfortunately, these methods are challenged by multiple factors, such as low accuracy, high complexity, etc. More importantly, these methods consider only historical trajectory data for prediction without taking into account time factors (i.e., the residence time of a user in a cell), which, however, is an important factor for the data traffic distribution in the wireless network.

In this paper, we study the user-level behavior patterns to predict the temporal and spatial data traffic distribution in the wireless network. To model a user profile, user's mobility and service pattern are investigated. An improved Bayesian trajectory prediction algorithm is proposed to predict the user's trajectory in the working and rest days, respectively. Also, the user residence time is considered to achieve more accurate trajectory prediction. User service pattern prediction is obtained by combining time feature and the mobility prediction results. Furthermore, instead of using virtual mobile data, our work is based on the real mobile data collected from the mobile networks, which can make our prediction results more meaningful. The contributions of this paper are summarized as below:

- (1) For user mobility prediction, an improved Bayesian prediction algorithm is conceived to predict the likelihood of the next location for users with considering the user residence time to achieve more accurate location prediction.
- (2) An innovative approach is designed to predict user service type and traffic by combining with time feature and location prediction results.
- (3) En-powered by user mobility and service patterns prediction, the effective prediction for temporal and spatial data traffic distribution is carried out.

2 User Profile Prediction Algorithm

The user profile prediction algorithm is divided into two parts: user's location prediction and service pattern prediction. For user's location prediction, the personal trajectory modeling can be conducted according to the users' historical trajectory. After modeling, when the system is inputted the continuous trajectory points of a user within a few hours, which cell the user will visit next time will be predicted as the output

results. For service pattern prediction, when the continuous service information for a period of time (for example, a few hours) is fed into the well trained model, the service type and the traffic used by the user at the next moment can be predicted as the output results. After the above two predictions are effectively combined, for a certain cell we can know how many users will enter and leave at the next moment, and how much data traffic they will bring in or take away. The real-time prediction of the users' mobility and service pattern will assist in the analysis of data traffic distribution in wireless networks at the near future moment.

The algorithm includes three parts (see Fig. 1): offline learning, online prediction and traffic statistics. In the offline learning stage, historical data is used to build prediction models. For online prediction, the real-time data in the process of user movement will be fed into the model that has been well trained in the offline learning stage to obtain the prediction results. Then, the data traffic at the next moment for a cell brought by each user can be summed up to obtain the total traffic. The other Key Performance Indicators (KPI) prediction (for example, PRB resource utilization) can be also made to understand the load condition of the network at the near future moment.

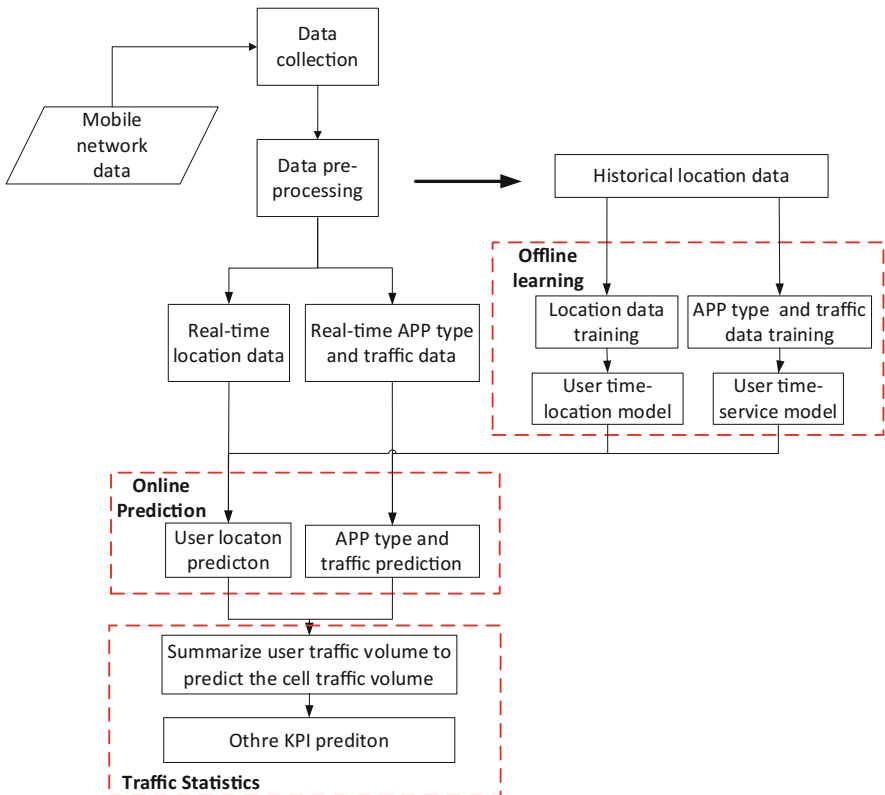


Fig. 1. The flowchart of prediction algorithm.

2.1 Mobility Prediction Model

In wireless network, mobile users need to have their existing connections transferred to a different access point (AP) to feed their highly mobile habits. With the movement of users, their connections are constantly being switched from one wireless AP to another. The user trajectory can be obtained by connecting the continuous wireless APs that have been accessed by users. So, the cell (or AP) granularity is taken as the smallest grid unit to analyze and predict the user's location. The trajectory T_1 can be represented as $\{n_1, n_2, n_3, n_6\}$ by network nodes.

A model based Bayesian theory towards this objective was studied in detail to make the user next location prediction. In this model, the user's previous moving trajectory has been known and is selected as historical trajectory. When the user's current travel trajectory is T_p , his destination n_j can be predicted by using Bayesian algorithm. And then the path with the highest probability of switching from T_p to a location l_n under the condition of destination n_j will be chosen and the next location of the trajectory l_n is the prediction location point. The formula is described as:

$$P(d \in n_j | T_p) = \frac{P(T_p | d \in n_j) \cdot P(d \in n_j)}{\sum_{k=1}^G P(T_p | d \in n_k) \cdot P(d \in n_k)}. \quad (1)$$

Where G represents the total number of destinations for a certain user. A location where a user resides for more than a certain time can be selected as the destination. The longer the residing time is, the more likely the location is to be a destination, but this method may miss the places where users only stay at a short time. For $P(d \in n_j)$, it is the priori probability of the destination n_j , it can be calculated as

$$P(d \in n_j) = S_{d \in n_j} / S_{total} \quad (2)$$

Where $S_{d \in n_j}$ is the total number of the historical trajectories that with the destination n_j . S_{total} is the total number of historical trajectories. Only a priori probability of the grid nodes that users have ever reached is not zero. That means, only the grid nodes that users have arrived before can be predicted as the user's next destination. It is important to note that the prior probability and the time are relevant. So, for better location prediction, the trajectory of the user's working day and the weekend is analyzed and counted, respectively. Where $P(T_p | d \in n_j)$ is the posteriori probability that indicates the probability of passing through the trajectory T_p when the destination is n_j . It can be calculated as:

$$P(T_p | d \in n_j) = \frac{S_{T_p, d \in n_j}}{S_{d \in n_j}} \quad (3)$$

Where $S_{T_p, d \in n_j}$ refers to the number of trajectories T_p when the destination is n_j . For the process of user trajectory prediction, the user's destination is predicted by real-time trajectory data, and then the cell with the highest handoff probability from the current

location to the predicted destination is selected. The next location of the trajectory T_p is the prediction point.

If the user is at the current location point n_c , the destination predicted is n_j , there may be different tracks from the current point to the destination, the location transfer probability from n_c to n_j is

$$P(n_c \rightarrow n_j) = P(n_c \rightarrow n_{c+1}) \cdot P(n_{c+1} \rightarrow n_{c+2}) \cdots P(n_{j-1} \rightarrow n_j) \quad (4)$$

The next location of the trajectory with $P(n_c \rightarrow n_j)$ maximum is the predicted location point.

According to the above model, the data of the rest day and the data of the working day are divided into two databases, which are trained and predicted respectively. The distinction between the working and the rest day can improve the accuracy of prediction model for it is in line with users' living habits.

2.2 Service Pattern Prediction Model

According to the burst and continuity of user's service pattern, we divide user's service type prediction into type analysis in the moving state and static state to make more accurate prediction. Service type modeling and prediction in motion state: the type of service used by the user is related to his trajectory. We decompose the user's probability of using certain application (APP) at some time and place to two important variables: APP type and location. After the user position switches from the current point to the next point, the APP that he uses may change. For example, when a user goes to the subway, he will turn on the video and watch the movie, and when he goes down the subway he will turn off the video. This event has a high occurring probability with a specific user. Define the state switching probability of using certain APP when user's location changes as:

$$P(q_r^a \rightarrow q_r^a | n_i \rightarrow n_{i+1}) = \frac{N(q_r^a \rightarrow q_r^a \& n_i \rightarrow n_{i+1})}{N(n_i \rightarrow n_{i+1})} \quad (5)$$

Where $N(q_r^a \rightarrow q_r^a \& n_i \rightarrow n_{i+1})$ refers to the number of events that the APP switches from $q_r^a \rightarrow q_r^a$ by a user when the location changes from $n_i \rightarrow n_{i+1}$. The next location n_{i+1} of user can be obtained by the previous trajectory prediction, and combine the current location n_i and the current application status q_r^a the number for handover $N(q_r^a \rightarrow q_r^a \& n_i \rightarrow n_{i+1})$ can get. Then the service type with the maximum $P(q_r^a \rightarrow q_r^a | n_i \rightarrow n_{i+1})$ will be chosen as the predicted service type.

In static state, according to the continuity of business, users' service type is related to the user's current location, next time period and the probability of a service type user will use in the next period of time. In this way, we can give the probability that the user will use a certain service type for the next period of time:

$$PA_{i,t,a} = \frac{N_{i,t,a}}{\sum_{a=1}^A N_{i,t,a}} \tag{6}$$

Among the formula, $N_{i,t,a}$ is the number of events that the service type a is used by the user at the location n_i and time t , and the A refers to the total number of user service types.

2.3 Traffic Statistics Model

The traffic of a certain wireless communication cell is determined by the usage of the users. All the traffic that users used in the cell needs to be counted to know the total traffic. As predicted above, whether a user will enter a cell can be measured by a probability value. Combined with the service type and traffic prediction, which APP is used by each user and how much traffic consumed is also a probability value. So the output of this module is also the probability of the total traffic volume of service APPs at the next time.

First, define the threshold for a cell traffic is V_{TH} , the probability of the total traffic is greater than V_{TH} will be calculated in this section. Define the traffic matrix V , in it the matrix v_{ij} indicates the traffic of user i uses the service type app j , and $i \leq m, j \leq n$. Matrix U is defined as a probability matrix for the user i to use a APP j , and $i \leq m, j \leq n$.

$$V = \begin{pmatrix} v_{11} & \cdots & v_{1n} \\ \cdots & \cdots & \cdots \\ v_{m1} & \cdots & v_{mn} \end{pmatrix} \tag{7}$$

$$U = \begin{pmatrix} u_{11} & \cdots & u_{m1} \\ \cdots & \cdots & \cdots \\ u_{1n} & \cdots & u_{mn} \end{pmatrix} \tag{8}$$

So, the traffic of user i is as follows:

$$\begin{pmatrix} V_1 \\ V_2 \\ \cdots \\ V_m \end{pmatrix} = V \cdot U = \sum_{j=1}^{A(j)} v_{ij} \cdot u_{ij} \tag{9}$$

The matrix E describes the number of users in the current cell and their state,

$$E = (e_1, e_2, e_3, \dots, e_{2^n}) \tag{10}$$

$$p = \begin{pmatrix} p_1 \\ p_2 \\ \cdots \\ p_m \end{pmatrix} \tag{11}$$

The probability matrix corresponding to the user state is defined as P , p_i is the probability of user i is in this cell. The total cell traffic S can be calculated as,

$$S = (V_1, V_2, \dots, V_m) \cdot E = (S_1, S_2, \dots, S_{2^n}) \quad (12)$$

In this case, the distribution probability of the total traffic volume of the cell is,

$$\begin{pmatrix} S_1 \\ S_2 \\ \dots \\ S_{2^n} \end{pmatrix} \rightarrow \begin{pmatrix} p_1 \\ p_2 \\ \dots \\ p_{2^n} \end{pmatrix} = p(e_i) \quad (13)$$

The probability that the total traffic volume is greater than V_{TH} in a cell is,

$$p(S > V_{TH}) = \sum_{i=1}^{2^n} p(e_i) p(S_i > V_{TH} | e_i) \quad (14)$$

3 Experiments Results

To verify the performance of our proposed scheme, we used a wireless trajectory dataset collected from Nanning to perform prediction. The format of input raw data is the detailed XDR. About 2300 mobile devices (users) active in 1600 wireless communication cells in NanNing urban area is chosen as test objects. The size of dataset collected is about 10G each day (including 265 million S1-U HyperText Transfer Protocol (HTTP) data and 223 million S1-U Mobility Management Entity (MME) data). History data of last 14 days is used to generate history training dataset. Every 15 min real time data in testing day is used as test dataset. For privacy security, data is encrypted to ensure that user's information is not directly involved in the study.

To estimate the overall accuracy of our prediction, we select the mobile users by their historical stay time. Similarly, the valid destinations are selected by historical stay number and time threshold. The track prediction result for 1 day (2018-01-23) is shown in Fig. 2. Blue line is the ratio of fully correct prediction. Red line is the ratio of next cell correct prediction. For most of the time, the track prediction accuracy is above 85%. Track prediction accuracy is a little lower in rush hour than other time. The mobile prediction is important to understand users' distribution in the network.

S1-U HTTP data is used for service pattern prediction. To estimate the overall accuracy of service type and traffic prediction, the proportion of correct service type and traffic prediction for all chosen mobile users is evaluated. The prediction result for one day (2018-01-23) is shown in Fig. 3. The overall service type and traffic prediction accuracy is above 70%, and the accuracy is a little lower in rush hour.

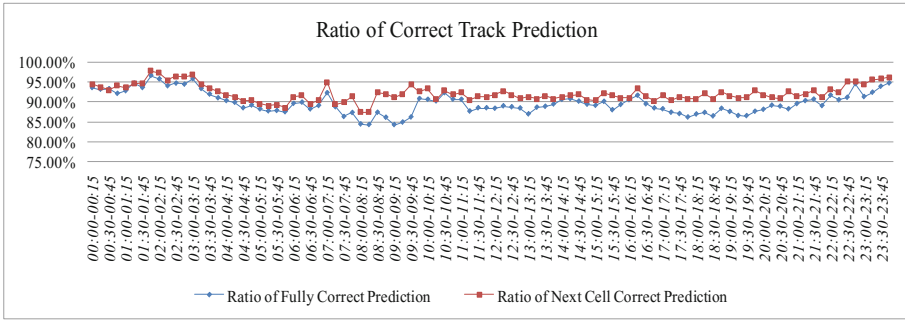


Fig. 2. Ratio of correct track prediction.

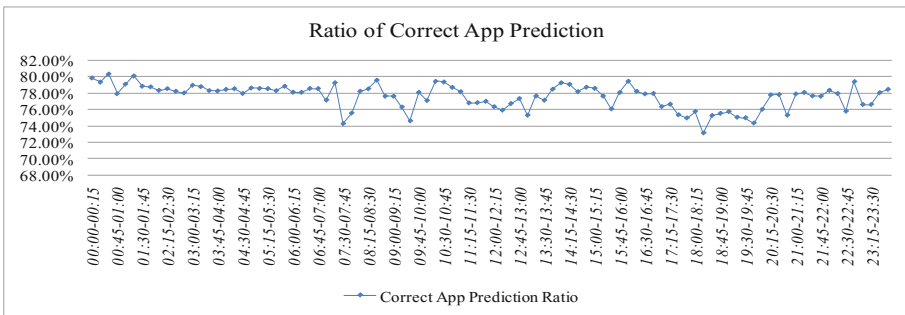


Fig. 3. Ratio of correct service type and traffic prediction.

4 Conclusions

In this study, we have proposed a new algorithm model for user profile prediction. The proposed scheme includes two sub-algorithms: Bayesian theory is used to model the user mobility pattern and an innovative method is designed to make service pattern prediction by combining the time and location features. The proposed strategy has low computational complexity. Field test results show that the accuracy and timeliness of the algorithm is of great significance to understand the temporal and spatial data traffic distribution in the wireless network. Also, the user profile prediction can help operators perceive the personalized service needs of mobile users so that more refined type of operation and maintenance can be carried out.

References

1. China Mobile Communications Corporation: Sustainability Report: Big Connectivity, New Future (2016)
2. Xu, L., Zhao, X., Luan, Y., et al.: User perception aware telecom data mining and network management for LTE/LTE-advanced networks. In: 4th International Conference on Signal and Information Processing, Networking and Computers, pp. 237–245. Springer, Qingdao (2018)

3. Xu, L., Luan, Y., Cheng, X., et al.: Telecom Big Data based user offloading self-optimisation in heterogeneous relay cellular systems. *Int. J. Distrib. Syst. Technol.* **8**(2), 27–46 (2017)
4. Song, L., Kotz, D., Jain, R., et al.: Evaluating next-cell predictors with extensive Wi-Fi mobility data. *IEEE Trans. Mob. Comput.* **5**(12), 1633–1649 (2006)
5. Tomar, R.S., Verma, S.: Trajectory prediction of lane changing vehicles using SVM. *Int. J. Veh. Saf.* **5**(4), 345–355 (2011)
6. Wang, J., Tang, J., Xu, Z., et al.: Spatiotemporal modeling and prediction in cellular networks: a Big Data enabled deep learning approach. In: *INFOCOM IEEE Conference on Computer Communications*, pp. 1–9. IEEE (2017)