Crowdsourced Mobility Prediction Based on Spatio-Temporal Contexts

Haitian Pang*, Peng Wang*, Lin Gao†, Ming Tang†, Jianwei Huang†, Lifeng Sun*
*Tsinghua National Laboratory for Information Science and Technology
Department of Computer Science and Technology, Tsinghua University
†Network Communications and Economics Lab (NCEL)
Department of Information Engineering, The Chinese University of Hong Kong

***Abstract*—Accurate mobility prediction is becoming increasingly important in human behavior research, mainly due to many location-based applications such as mobile social networks and mobile advertisements. In this work, we propose a new crowdsourced human mobility prediction model for public regions. We first analyze human trajectories collected through a cluster of densely deployed Wi-Fi access points (AP) in a shopping mall, and then characterize the close relationship between the human mobility patterns and the spatio-temporal contexts. Based on the distinct features of human trajectories in different types of public regions, we further propose a Markov-based crowdsourced mobility prediction method utilizing spatio-temporal contexts. We evaluate the performance of the proposed method using real traces, and show that our method is 28% more accurate in predicting human location transitions and incurs 14% smaller error in stay time prediction than the baseline methods.**

I. Introduction

Obtaining accurate human mobility information is important in many fields, such as the city planning, location based social networking, route optimization [1], smartphone collaboration [2] and mobile advertisement. The mobility prediction in public regions (such as shopping malls, airports, and conference venues) is especially important because the traces of a large number of visitors can provide rich information regarding the human mobility patterns. In this paper, we focus on the human mobility prediction in the public regions, using data traces recorded by the wireless local area networks (WLANs).

There are several popular mobility prediction approaches in the literature. One approach is to predict based on location information provided by cellular networks (e.g., such as the social based model [3], Markovian model [4], [5], and hierarchical model [6]). The cellular based prediction often has difficulty in terms of differentiating human mobilities among relative small areas due to its coarse localization granularity. Another approach is based on the mobile phone's GPS signals (e.g., [7]–[10]). These studies often focused on predicting users' mobility behaviors between their private locations such as "home" and "office". Although GPS can capture human outdoor position changes seamlessly and continuously, it often fails to capture human indoor movements due to weak satellite signals. To overcome the disadvantages of cellular and GPS based predictions discussed above, in the paper we focus on using traces from WLANs (i.e., Wi-Fi access points) to model human indoor mobility in the public area.

There are several existing studies on the mobility prediction based on WLAN data traces (e.g., [11]–[13]). Hsu *et al.* in [11] proposed the first WLAN-based mobility model that captures heterogeneous human behaviors in both spatial and temporal domains. Zhuang *et al.* in [12] focused on the study of social relationship and geographic coincidence prediction. Manweiler *et al.* in [13] proposed a method to predict a human's length of stay at a specific Wi-Fi hotspot. The above literature did not discuss the human mobility issues in public regions.

Some prior literatures worked on predicting how a fixed group of humans move across several frequently visited places such as "home" and "office" (e.g., [7]–[10], [14], [15]). However, the methods they proposed cannot be directly applied to predict human mobility in public regions (such as shopping malls and airports), which involves a large number of users with relatively irregular movement patterns. More specifically, comparing with humans whose mobility partterns are derived from sufficient data traces related to a limited number of popular locations [8], humans move more randomly over a large number of locations in public regions and leave shorter traces. Therefore, it is more challenging to build an accurate mobility model for each single user in a public area based on existing methods due to lack of per user data.

To deal with the above challenges, we utilize crowdsourced historical movement data, which consist of short trajectories of many different users, to reflect and predict users' mobility tendency in specific regions. Hence our focus of data collection and analysis is *location-oriented*, instead of *user-oriented* in previous literature (e.g., [11], [16]). Note that there is also some existing *location-oriented* work. For example, Balachandran *et al.* in [17] conducted an empirical study on the human behavior and network performance in a public-area wireless network. To enable fast handoff among Wi-Fi access points, Pack *et al.* in [18] proposed a mobility prediction based method, which focuses on Wi-Fi access points with overlapped coverages. Different with these work, we will analyze various

The work of H. Pang, P. Wang, and L. Sun is supported by the NSFC under Grant No.61272231,61133008, 61502264 and Beijing Key Laboratory of Networked Multimedia. This work of L. Gao, M. Tang, and J. Huang is supported by the General Research Funds (Project Number CUHK 412713, 14202814, and 14206315) established under the University Grant Committee of the Hong Kong Special Administrative Region, China.

978-1-4799-6664-6/16/$31.00 ©2016 IEEE

crowd behavior patterns at different locations, and embed them into our prediction model.

More specifically, we utilize the *spatio-temporal context information* to characterize the crowd mobility preferences at different locations. We will investigate the following questions in this paper: *(1) How do temporal and spatial contexts affect users' mobility preferences? (2) How can we utilize these contexts to improve the prediction performance?* To answer the first question, we extract spatio-temporal contextual features based on the group mobility history generated by a large amount of users, to characterize users' visit probabilities at specific locations. To answer the second question, we embed the spatio-temporal contexts into the existing Markov mobility prediction model (e.g., [19], [20]), and derive a new crowdsourced mobility prediction model. We evaluate the performance of the proposed model by using real traces, and show that our new model outperforms the widely used Markov model, with an average 28% of prediction accuracy improvement. The proposed method also outperforms the Markov model in stay time prediction with 14% of error reduction.

In Section II, we will explain preliminaries on *contexts*. In Section III, we propose a crowdsourced mobility prediction model based on the result in Section II. In Section IV, we present the evaluation of our proposed mobility prediction model using real traces. Finally we conclude in Section V.

II. Preliminaries on Contexts

We collected the mobility trace from NextWiFi [21], a recently launched commercial wireless service provider in Asia. To understand the relationship between physical contexts and mobility patterns, we analyzed the data collected at 30 Wi-Fi APs deployed in a four-floor shopping mall (public region). Fig. 1 illustrates the AP locations on each floor. We classify the 30 AP locations into 4 function zones: restaurant, women's clothing, men's clothing, and elevator. Fig. 1 shows that APs are not deployed in a regular pattern, which makes the human mobility prediction difficult.

The NextWiFi dataset contains 76,933 unique human trajectories over the period of one month. Fig. 2 shows the cumulative distribution function (CDF) of the trajectory lengths of all users in the dataset, and the average length of the trajectories equals 9.1. This significantly differs from WLAN human mobility traces reported in some previous work. For example, the dataset in [20] contains Wi-Fi traces collected in a public campus and is widely used in many previous work. Users in the dataset tend to visit a much larger number of APs (e.g., the median number is 494 in [20]).

Next we discuss how to extract context information and use it accordingly in the prediction of human mobility based on these real traces. In mobility prediction, a *context* refers to the location-based crowd visit information which can be used to characterize crowd movement. Specific contexts include the function zone of a location, the connection frequency of a function zone, the long stay ratio in a location, and transition preferences between locations, etc. Several temporal-spatial contexts are now described in more details.

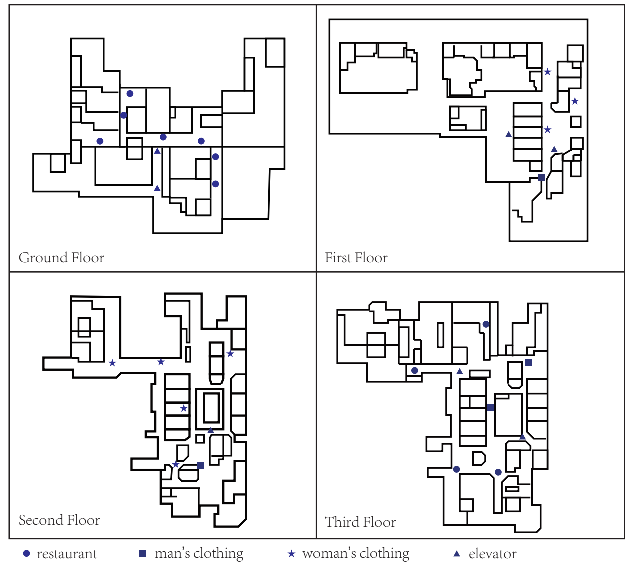

Fig. 1. Map and floor layouts of the shopping mall.

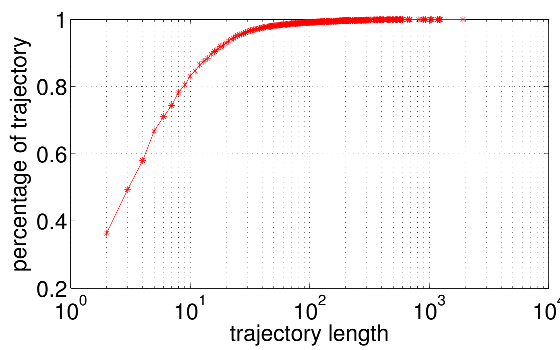

Fig. 2. CDF of trajectory length.

A. Connection frequency

Connection frequency is defined as the number of Wi-Fi connections measured at a specific function zone during a particular time period. Clearly a higher connection frequency corresponds to a high visiting preference from users. In the following discussions, we consider a time-slotted system, with each time slot corresponding to 1 minute. For example, considering the fact that the shopping mall opens at 8:30 a.m. and closes at 11:10 p.m., then a day includes 880 time slots. Fig. 3 illustrates such a context, where the x-axis represents the time, and the y-axis represents the connection frequency at all Wi-Fi APs in a specific function zone. We can see that the connection frequency has obvious peaks during lunch and dinner periods in the restaurant function zone. However, the connection frequency stays relatively high from noon to evening in the clothing zones (for both women and men). Elevator zone gets visits first at the open time of the

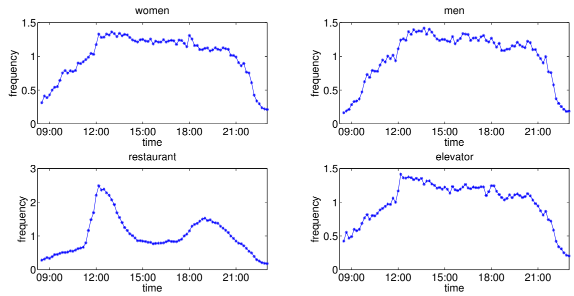

Fig. 3. Connection frequencies in different function zones. The x-axis depicts the time from 8:30 to 23:10, and y-axis is the connection frequencies.

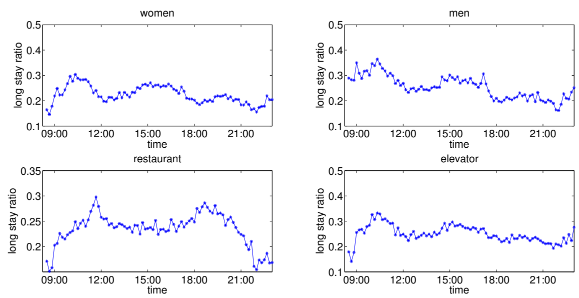

Fig. 4. Long stay ratio of function zones in a day. The x-axis is the time from 8:30 to 23:10 and y-axis is long stay ratio.

shopping mall, because users swarm into elevators to reach other shopping zones.

We use matrix $\boldsymbol{V}_c = [V_c(t,f)]$ to denote the connection frequencies in different time slots and function zones. Here subscript c represents "connection", t is the time index, and f is the function zone index. We divide one day into a set $\mathcal{T} = \{1, 2, ..., T\}$ of time slots, and denote the set of locations as $\mathcal{L} = \{0, 1, ..., 30\}$. Here 0 represents the area that is not covered by any Wi-Fi AP. Each location belongs to one of the four function zones in set $\mathcal{F} = \{1, 2, 3, 4\}$. We denote $n(t, f)$ as the number of connections during time slot t at function zone f. Hence the connection frequency of function zone f at time t can be calculated as

$$V_{\rm c}(t,f) = \frac{n(t,f)}{\frac{1}{T}\sum_{t'\in\mathcal{T}} n(t',f)}.$$

B. Long stay ratio

We define *long stay* as a connection with a duration of no smaller than 10 minutes, and the *long stay ratio* as the percentage of long stays among all recorded Wi-Fi connections. Fig. 4 shows different long stay ratios in different function zones. The reason to consider the long stay ratio is to differentiate intentional and unintentional human visits, under the assumption that long stays are due to intentional decisions. Fig. 4 clearly shows that the long stay ratios reach peaks during lunch and dinner periods in the restaurant zone. For the other function zones, the peaks appear in the morning, afternoon and at night. This is different from the connection frequency in Fig. 3, which only reflects the visiting preferences without considering the factor of staying time.

We use matrix $\boldsymbol{V}_s = [V_{\rm s}(t,f)]$ to denote the long stay ratio at different time slots and different function zones. Here subscript s denotes the stay ratio, t denotes the time index, and f denotes the function zone index. The long stay ratio of function zone f at time t can be calculated as

$$V_{\rm s}(t,f) = \frac{n_{\rm long}(t,f)}{n(t,f)},$$

where $n_{\rm long}(t,f)$ denotes number of long stays during time slot t at function zone f.

C. Transition Preference

Transition preference reflects how users move across different locations, which is affected by the topology of functional deployments, and physical proximity of them. Fig. 5 illustrates how a user changes locations over time. We can characterize a user's trajectory as series of connection transitions among different locations. In Fig. 5, the trajectory is $\{1, 3\}$ and we have omitted the location 0 that only means a location that is not covered by any Wi-Fi APs.

Mathematically, we denote $l \in \mathcal{L}$ as the index of locations. Hence we consider Transition Markov model and construct the transition matrix $\boldsymbol{T}_{\rm l} = [T_{\rm l}(l,l)]$, illustrating the transition probabilities among different locations. Suppose the future location $l_{\rm future} = l'$ and current location $l_{\rm now} = l$, we can obtain $T_{\rm l}$ by

$$T_{\rm l}(l,l') = P(l_{\rm future} = l' | l_{\rm now} = l), \tag{1}$$

where $P(l_{\rm future} = l' | l_{\rm now} = l)$ can be calculated as

$$\frac{m\{l_{\rm future} = l', l_{\rm now} = l\}}{\sum_{l'} m\{l_{\rm future} = l', l_{\rm now} = l\}},$$

where $m\{l_{\rm future} = l', l_{\rm now} = l\}$ denotes the number of transitions from l to l' that we observe in the dataset of all users.

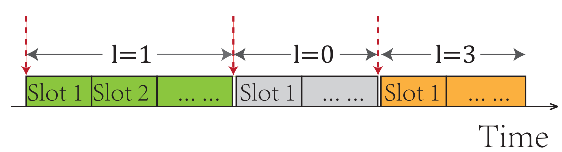

Fig. 5. An example of a user's time-slotted trace.

D. Transition impact on stay time

We notice that the location of a user's Wi-Fi connection will affect the time that he stays in this location. We define a matrix $\boldsymbol{I}_{\rm f} = [I_{\rm f}(f,f)]$ to reflect such an impact. The normalized value of each element of $\boldsymbol{I}_{\rm f}$ is given in Table II.

$$I_{\rm f}(f,f') \triangleq \frac{mean\_stay\_time(f'|f)}{mean\_stay\_time(f')}$$

where $mean\_stay\_time(f')$ is the mean stay time in function zone f', and $mean\_stay\_time(f'|f)$ is the mean stay time of users in function zone $f^{'}$ conditional on the fact that the user was in function zone f in the previous connection.

In particular, $I_{\mathrm{f}}(f, f^{'}) > 1$ means a positive influence of zone f on zone $f^{'}$, i.e., being in function zone f before is likely to lead to a longer stay at function zone $f^{'}$. The magnitude of $I_{\mathrm{f}}(f, f^{'})$ represents the intensity of the impact. For example, the last row of Table II shows that users coming from restaurant zone tend to spend less time in clothing zones, maybe because most users prefer to walk around after meals instead of going shopping directly. On the other hand, the second to last row shows that a user is likely to spend more time in clothing zones if he has just spent time in an elevator zone, because users usually take the elevators before changing shopping zones.

TABLE I

IMPACT OF PREVIOUS FUNCTION ZONE ON CURRENT STAY TIME. THE COLUMN REPRESENTS THE CURRENT FUNCTION ZONE, AND THE ROW REPRESENTS THE PREVIOUS FUNCTION ZONE.

	men's	women's	elevator	restaurant
men's clothing	1.005	1.295	1.246	1.086
women's clothing	1.060	1.041	1.045	0.969
elevator	1.133	1.136	1.002	0.982
restaurant	0.900	0.821	0.869	1.036

E. Mean Stay Duration

We define *Mean Stay Duration* of a user as his average stay length at all APs. Fig. 6 illustrates the CDF of all users' mean stay lengths. We denote users in the shopping mall Wi-Fi network as $\mathcal{U} = \{1, 2, ..., U\}$. For convenience, we use $S_{\mathrm{m}}(u) \triangleq \log(1 + mean\_time(u))$ to characterize a user u's stay time preference. Later we will use matrix $\boldsymbol{S}_{\mathrm{m}}$ to predict a user's stay time at a location.

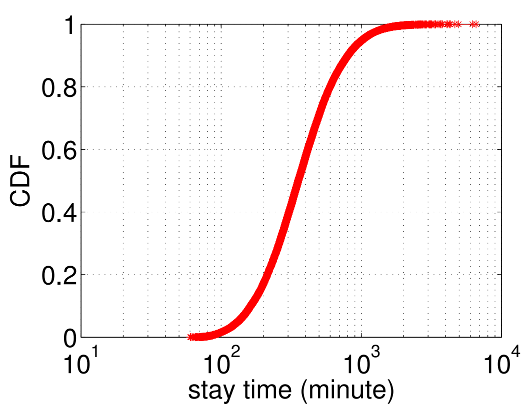

Fig. 6. Cumulative distribution of users' mean stay time.

To summarize, the contexts *Connection frequency* and *Long stay ratio* involve the time and function zone. The contexts *Transition preference* and *Transition impact on stay time* involve locations or the function zone of locations. Finally, the context *Mean stay duration* reflects the individual stay preference.

III. PREDICTION BASED ON SPATIO-TEMPORAL CONTEXTS

In this section, we will propose a mobility prediction model incorporating the context information discussed in Section II, namely, $\boldsymbol{V}_{\mathrm{c}}$, $\boldsymbol{V}_{\mathrm{s}}$, $\boldsymbol{T}_{\mathrm{l}}$, $\boldsymbol{I}_{\mathrm{f}}$, and $\boldsymbol{S}_{\mathrm{m}}$. We first introduce the Markov predictor model given in [19], [20], based on which we will propose the crowdsourced method that incorporates contexts previously defined.

A. Individual Markov Mobility Predictor

We first review the Markov mobility predictor introduced in the literature, and use it to predict the location change between successive time slots in our model. Baummann *et al.* in [8] demonstrated that the Markov predictor performs well on Wi-Fi mobility data in terms of accuracy. Song *et al.* in [20] also showed that the low-order Markov predictors perform better than the high-order Markov predictors. Motivated by these earlier studies, we choose Markov predictor as the baseline method in this paper[1].

A user u's path in a day (PPD_u) can be represented by a vector of 880 elements, i.e.,

$$PPD_u = \{l_u[t], t = 1, ..., 880\},$$

where $l_u[t]$ represents the AP ID (i.e., location) associated with user u at time slot t.

Based on PPD_u, we can train the transition matrix $\boldsymbol{M}_u = [M_u(l, l^{'})]$ of user u

$$M_u(l, l^{'}) = P(l[t+1] = l^{'} | l[t] = l), \tag{2}$$

where $P(l[t+1] = l^{'} | l[t] = l)$ stands for the first-order transition probability computed as follows:

$$\frac{m\{l_u[t+1] = l^{'}, l_u[t] = l\}}{\sum_{\bar{l}} m\{l_u[t+1] = \bar{l}, l_u[t] = l\}},$$

and $m\{l_u[t+1] = l^{'}, l_u[t] = l\}$ denotes the number of transitions from location l to location $l^{'}$ in PPD_u. Note that $M_u(l, l^{'})$ is the transition between two successive time slots, while $T_{\mathrm{l}}(l, l^{'})$ defined in Section II-C is the transition between two successive connections.

B. Crowdsourced Markov Mobility Predictor

In public regions, having an accurate understanding of crowd mobility preference is important for the mobility prediction. As the trajectory of a single user tends to be short (e.g., the trajectory length in Table I equals 4), we employ a crowdsourced method to train the transition matrix $\boldsymbol{M}$. More specifically, we calculate the transition matrix with all users' PPDs. This will capture the crowd mobility patterns in specific regions, regardless of the short historical trajectory of each single user. While losing the possibility of analysis at an individual level, this crowdsourced method allows us to build a more accurate model at the crowd level. The generation of crowd transition matrix $\boldsymbol{M}$ of the crowd is similar to the calculation of $\boldsymbol{M}_u$ in Section III-A.

[1]We also reach the similar conclusion in the experiment, as second-order predictor can only improve the first-order predictor by 1% in terms of prediction accuracy. Hence we focus on first-order Markov predictor.

C. Spatio-Temporal Context-Based Markov model

Markov Predictor has its advantage in time sequence prediction, but it fails to represent contexts such as connection frequencies and long stay ratio. As these contexts reflect users' mobility preferences about locations and times, we incorporate the parameters of temporal and spatial contexts in Section II into the prediction model. We define the contextual factor $\mathbf{\Theta} = (\Theta_u, \forall u)$ to capture the user's stay duration preferences *when the user appears*. We also define the contextual factor Φ to capture the user's transition preferences *when the user does not appear*.

1) Prediction when the user appears : When a user u stays at the location l, and given that the start time of the connection is $t_{\text{start}} = t_0$ and the location of previous connection is $l_{\text{pre}} = l^{'}$, the probability that the user stays at the same location in the next time slot is affected by contexts concerning stay length. Recall that matrix $\boldsymbol{V}_{\text{s}}$ represents the long stay ratio at different locations, matrix $\boldsymbol{I}_{\text{f}}$ reflects the transition impact on stay length, and $\boldsymbol{S}_{\text{m}}$ indicates users' preferences of stay lengths[2]. We denote the function zone of location l as $f(l)$. Using the above notations, the contextual factor Θ_{u} of user u can be formalized as

$$\begin{aligned} &\Theta_{\text{u}}(l[t+1] = l | l[t] = l, l_{\text{pre}} = l^{'}, t_{\text{start}} = t_0) \\ &\triangleq V_{\text{s}}(t_0, f(l)) I_{\text{f}}(f(l^{'}), f(l)) S_{\text{m}}(u). \end{aligned} \tag{3}$$

With this formalization, a user u tends to stay longer if $\Theta_u > 1$.

2) Prediction when the user does not appear: When the current location is 0, and given the location before transition $l_{\text{pre}} = l^{'}$ and the user is in the process of a transition between locations, the mobility can be affected by contexts concerning location transitions. On one hand, crowd location transition preference is affected by the transition matrix $\boldsymbol{T}_{\text{l}}$; on the other hand, the probability that a location gets visited is determined by $\boldsymbol{V}_{\text{c}}$. We define the contextual factor Φ in this scenario as

$$\begin{aligned} &\Phi(l[t+1] = l | l[t] = 0, l_{\text{pre}} = l^{'}) \\ &\triangleq T_{\text{l}}(f(l^{'}), f(l)) V_{\text{c}}(t, f(l^{'})). \end{aligned} \tag{4}$$

The larger value of Φ indicates a larger probability of transition to location l in the next time slot.

Based on the definition of the above $\mathbf{\Theta}$ and Φ, we finally achieve a probabilistic matrix $\boldsymbol{P} = [P(l[t+1] | l[t], l_{\text{pre}}, t_{\text{start}})]$ by embedding contextual factors into first-order Markov Predictor $\boldsymbol{M}$.

$$\begin{aligned} &P(l[t+1] = l_3 | l[t] = l_2, l_{\text{pre}} = l_1, t_{\text{start}} = t_0) = \\ &\begin{cases} M(l_2, l_2) \cdot \Theta_{\text{u}}(l_2 | l_2, l_1, t_0), & l[t] = l[t+1], \\ M(0, l_3) \cdot \Phi(l_3 | l[t] = 0, l_{\text{pre}} = l_1), & l[t] = 0, \\ M(l_3, l_2) \cdot 1, & \textit{otherwise}. \end{cases} \end{aligned} \tag{5}$$

Based on the calculation in (5), we need to further normalize $P(l[t+1] | l[t], l_{\text{pre}}, t_{\text{start}})$ to guarantee that the matrix is a probabilitic matrix, which depends on the current location, the previous location (before transition), and other contexts. In this way, contextual parameters are incorporated into the crowdsourced Markov mobility model in our proposed prediction method. We can also utilize the proposed mobility model to predict how long the user will stay at the current location. The performance of the proposed method will be evaluated in predicting mobility transition and the corresponding stay time.

[2]For simplicity, we assume the process is memoryless.

IV. EXPERIMENT EVALUATION

We divide the entire dataset into two parts: the training set and the test set. We randomly select a part of users' trajectories as the test set, and train the crowdsourced mobility model with the rest of trajectories.

We compare our proposed temporal-spatial crowdsourced (TSC) method with two benchmark methods discussed in Section III: CROWDSOURCED method and INDIVIDUAL method. We utilize the crowd's data trace to generate the model of CROWDSOURCED method with first-order Markov, and use data traces of individual users to generate the model of INDIVIDUAL method. Suppose that the current time slot is t_0, and the location after t time slots is $PPD[t_0 + t]$. We compare the predicted $PPD[t_0 + t]$ of the two Markov models with the ground truth. In Fig. 7, we compare the prediction accuracy of the three methods. The x-axis denotes the number of time slots in our prediction. The y-axis denotes the percentage that we make correct predictions. Fig. 7 shows that when the gap t increases, the accuracies of all three methods decrease. However, the average accuracy of our method is 28% higher than that of the CROWDSOURCED method. The INDIVIDUAL method performs the worst in terms of accuracy.

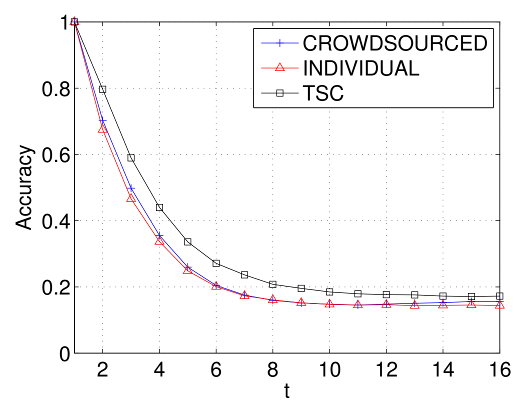

Fig. 7. Accuracy of predicting $PPD[t_0 + t]$

Fig. 8 shows the cumulative distribution function (CDF) of accurate predictions for the next 40 time slots ($t = [1 : 40]$). The x-axis has the same meaning as in Fig. 7, and the y-axis is the cumulative distribution. As shown in Fig. 8, our proposed method performs the best in terms of accuracy. The average number of correct predictions of our method 15% more than that of the INDIVIDUAL method.

We further evaluate the performance of the proposed prediction method in terms of predicting the stay time at a

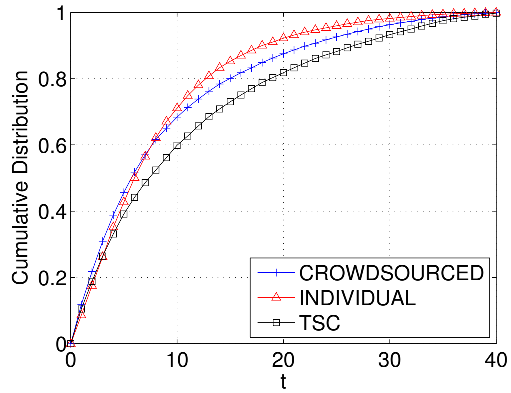

Fig. 8. CDF of accurate prediction number

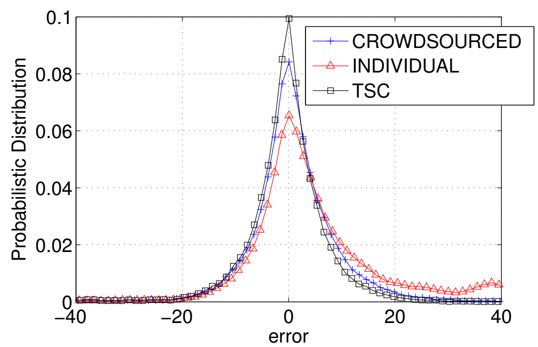

Fig. 9. The PMF of stay time prediction error

user's current location, which is a very challenging practical problem [8]. We define *predicted stay time* R_p and *actual stay time* R_r, and define *error of stay time prediction* as $E_p = R_p - R_r$. A smaller value of E_p indicates a more accurate prediction. In an evaluation on a large-scale test set with $R_r \leq 40\ slots$, we obtain the distributions of E_p returned by three methods as shown in Fig. 9. In Fig. 9, the x-axis is the prediction error, and the y-axis is the probability mass function. The prediction errors of all methods empirically follow similar distributions with mean zero. However, our method outperforms CROWDSOURCED and INDIVIDUAL methods due to a smaller variance. The average error of our method is 14% smaller than that of CROWDSOURCED method in stay time prediction.

V. CONCLUSIONS

In this paper, we studied human mobility patterns in public regions. We conducted an empirical study on the human mobility patterns based on data recorded by Wi-Fi APs, and extracted some important temporal and spatial contexts from the dataset. Considering the challenge of per user mobility prediction based on short movement trajectories in public regions, we proposed a crowdsourced method that aggregated a large amount of information of the entire crowd into a single location transition model. We further applied the contextual features to improve the performance of Markov model, and proposed a spatio-temporal context-based prediction method. Our method can be used to predict the transitions among locations as well as stay time in the current location. Performance evaluation based on the realistic data showed that our proposed method achieves significantly higher prediction accuracies over two benchmark algorithms.

REFERENCES

[1] A. Z. M. Shahriar, M. Atiquzzaman, and W. Ivancic, "Route optimization in network mobility: Solutions, classification, comparison, and future research directions," *Communications Surveys & Tutorials, IEEE.*

[2] L. Duan, T. Kubo, K. Sugiyama, J. Huang, T. Hasegawa, and J. Walrand, "Incentive mechanisms for smartphone collaboration in data acquisition and distributed computing," in *IEEE INFOCOM*. IEEE, 2012.

[3] D. Zhang, D. Zhang, H. Xiong, L. T. Yang, and V. Gauthier, "Nextcell: Predicting location using social interplay from cell phone traces," *IEEE Transactions on Computers*, 2015.

[4] G. Y. Liu and G. Q. Maguire Jr, "A predictive mobility management algorithm for wireless mobile computing and communications," in *IEEE International Conference on Universal Personal Communications*. IEEE, 1995.

[5] J. B. Othman, H. Castel-Taleb, and L. Mokdad, "Performance evaluation of mobile networks based on stochastic ordering of markov chains," *Studia Informatica Universalis*, 2006.

[6] T. Liu, P. Bahl, and I. Chlamtac, "Mobility modeling, location tracking, and trajectory prediction in wireless atm networks," *IEEE JSAC*, 1998.

[7] H. Gao, J. Tang, and H. Liu, "Mobile location prediction in spatio-temporal context," in *Nokia mobile data challenge workshop*, 2012.

[8] P. Baumann, W. Kleiminger, and S. Santini, "How long are you staying?: predicting residence time from human mobility traces," in *Proceedings of the 19th annual international conference on Mobile computing & networking*. ACM, 2013.

[9] T. M. T. Do and D. Gatica-Perez, "Contextual conditional models for smartphone-based human mobility prediction," in *Proceedings of the 2012 ACM conference on ubiquitous computing*. ACM, 2012.

[10] S. Scellato, M. Musolesi, C. Mascolo, V. Latora, and A. T. Campbell, "Nextplace: a spatio-temporal prediction framework for pervasive systems," in *Pervasive Computing*. Springer, 2011.

[11] W.-j. Hsu, T. Spyropoulos, K. Psounis, and A. Helmy, "Modeling time-variant user mobility in wireless mobile networks," in *IEEE INFOCOM*. IEEE, 2007.

[12] H. Zhuang, A. Chin, S. Wu, W. Wang, X. Wang, and J. Tang, "Inferring geographic coincidence in ephemeral social networks," in *Machine Learning and Knowledge Discovery in Databases*. Springer, 2012.

[13] J. Manweiler, N. Santhapuri, R. R. Choudhury, and S. Nelakuditi, "Predicting length of stay at wifi hotspots," in *Proc. of IEEE INFOCOM*. IEEE, 2013.

[14] Y. Chon, H. Shin, E. Talipov, and H. Cha, "Evaluating mobility models for temporal prediction with high-granularity mobility data," in *Percom*. IEEE, 2012.

[15] J. Ghosh, M. J. Beal, H. Q. Ngo, and C. Qiao, "On profiling mobility and predicting locations of wireless users," in *Proceedings of the 2nd international workshop on Multi-hop ad hoc networks: from theory to reality*. ACM, 2006.

[16] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: user movement in location-based social networks," in *Proc. of ACM SIGKDD*. ACM, 2011.

[17] A. Balachandran, G. M. Voelker, P. Bahl, and P. V. Rangan, "Characterizing user behavior and network performance in a public wireless lan," in *ACM SIGMETRICS Performance Evaluation Review*. ACM, 2002.

[18] S. Pack and Y. Choi, "Fast handoff scheme based on mobility prediction in public wireless lan systems," *IEE Proceedings-Communications*, 2004.

[19] J.-K. Lee and J. C. Hou, "Modeling steady-state and transient behaviors of user mobility: formulation, analysis, and application," in *Proc. of the 7th ACM international symposium on Mobile ad hoc networking and computing*. ACM, 2006.

[20] L. Song, D. Kotz, R. Jain, and X. He, "Evaluating location predictors with extensive wi-fi mobility data," in *IEEE INFOCOM*. IEEE, 2004.

[21] http://www.nextwifi.cn/wifind/.