

Feature Tendency Based Location Prediction in LBSNs

Zi Xing, Hui Tian, Tu Chen, Jing Zhang

State Key Laboratory of Networking and Switching Technology

Beijing University of Posts and Telecommunications, Beijing 100876, China

xingzi_cuc@163.com, tianhui@bupt.edu.cn, chentu_1991@163.com, jingzhang1919@bupt.edu.cn

Abstract—The development of location-based social networks (LBSNs) has brought in massive users' mobility data, providing an unprecedented opportunity to study human mobile behavior. However, the existing location prediction methods suffer from incompleteness of mobility data and disability of selecting the effective feature. To better understand and describe mobile behavior, users' multi-features are induced and reorganized comprehensively in this paper. Meanwhile, we propose a novel location prediction model, *Powerful Feature Tendency Based Selection (PFTS)*, which is universal for all datasets with diverse features. Furthermore, the proposed model can self-adaptively select the feature with the most powerful tendency by comparing variance of visiting probability distribution. Based on the chosen features, our model automatically employs corresponding method to predict locations for various scenes. Using two real-world LBSNs datasets, the experiment results validate that the *PFTS* significantly outperforms state-of-the-art approaches in terms of prediction accuracy. The results also suggest that the feature tendency is efficient to capture users' mobile behavior for boosting prediction performances.

Index Terms—Location-based social networks, Location prediction, Mobile behavior, Self-adaptively select, Feature tendency

I. INTRODUCTION

The proliferation of mobile devices provides the opportunity to gain insights on users' visiting behavior using mobile data, such as GPS [1], wifi signals [2], small card transactions [3], or location check-ins from location-based social networks (LBSNs). With easy accessibility and good intuitiveness, check-in data in online LBSNs, including user data, location data, corresponding time stamps, as well as users' friendship, is widely utilized to research human mobility. As a focus of LBSNs, location prediction has attracted more enthusiasm, which is to predict users' future locations based on historical position information, as shown in Fig. 1. The predictability of human mobility with a limit of 93% demonstrates that human mobility is predictable [4]. Over the past years, location prediction has already benefited many areas in our life, such as targeted advertising [5], urban planning [6], accidents avoiding and disease spread preventing [7]. Moreover, many research groups have investigated the enhancement of the fifth-generation mobile network (5G) brought about by user location prediction [8], including user trajectory [9], hot-spots [10], motion scope [11], and so on to capture user behavioral preference. Given users' potential visiting locations, radio

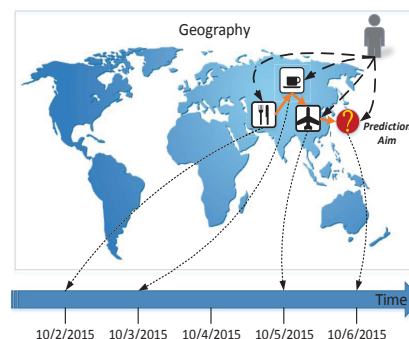


Fig. 1: Location Prediction of Human Mobility

resource managers can better pre-allocate resource to satisfy more users and reduce energy-consumption. Therefore, an accurate prediction of user mobility is hence essential for guiding various industries to deploy works in advance and promoting business benefits.

Recently, a number of studies have explored location prediction on LBSNs [12]. Chang et al. [13] merely consider spatial information of users' historical check-ins for location prediction. In contrast, researchers [14], [15] only introduce temporal information on LBSNs, including the temporal periodicity in terms of days' same hour and weeks' same day, to predict users' next locations. While, Cho et al. [16] suggest that people commute to and from work at roughly the same time during workdays, as opposed to weekends when people's travel and schedules are less predictable. Not only does the similarity of workdays prove the fine-granularity research on everyday of a week to be unnecessary, but also the actual experiments show low prediction accuracies on this research because of the sparsity data. Differing from prediction models in [13]–[16] for LBSNs, Gao et al. [17], [18] propose Gaussian mixture model (GMM) to combine part spatial and temporal features for location prediction, unfortunately showing boundedness due to the requirement of concurrence of spatial and temporal information. Furthermore, traditional algorithms [13]–[19] of location prediction on LBSNs seldom concern the feature tendency of user mobility.

In this paper, in order to eliminate the limitation of concurrence of spatio-temporal information, conquer unreasonable division of temporal periodicity, and focus on a user's

personality of behavioral preferences, we inevitably need to consider two aspects: (1) **mobile feature category**: in terms of user behavior; (2) **feature tendency**: behavioral tendency for various locations in the view of mobile features. Owing to the fact that applying different features to predict locations leads to different accuracies, we propose an innovative model named *Powerful Feature Tendency Based Selection (PFTS)*, to boost prediction performance. The main contributions of our work can be summarized as following:

- Based on the behavior analysis, we comprehensively reorganize users' mobile features with distinct feature parameters on LBSNs, which is convenient to describe mobile characters for modeling location prediction.
- By clustering days of a week into workdays and weekends to introduce temporal patterns, we overcome the inaccuracy of location prediction due to the traditional fine-granularity division.
- We propose an universal location prediction model, named *PFTS*, which considers all mobile features after inducing and integrating user behavior. A self-adaptive selection, which uses variance of visiting's probability distribution to find the feature with the most powerful tendency, is conducted for various check-ins. The proposed model automatically employs corresponding analytical method to predict locations, based on the chosen feature.
- We investigate check-in data from real LBSNs datasets (Gowalla [19] and Brightkite [18]). The experimental results show that our model significantly improves prediction accuracy comparing with state-of-the-art methods.

II. ANALYZING USER MOBILE BEHAVIOR

In order to predict locations, we analyze mobile features from historical check-ins through spatial and temporal perspectives to capture user behavior. All features are discussed below:

1) Spatial Feature:

a) **Spatial Independence Feature (SIF)**: Based on distinctive check-in frequencies at different locations, we can find users' visiting preference in locations. Chang et al. [13] have proved that the check-in frequency of a user u , only considering the location set, is an efficient predictor, so we define SIF as the check-in frequency of u 's visited locations.

b) **Spatial Markov Feature (SMF)**: In daily life, some people tend to check in at a particular place after another, e.g., a user usually checks in at two locations consecutively: the coffee house and the company. Therefore, by Markovian character of human mobility, we can define SMF as the check-in frequency of u 's next location l via the latest check-in location l' .

2) **Temporal Feature**: Check-in temporal information, related to many cyclic features, can be implied by time $t \in T$ (e.g., 2010-05-17 08:30:25) to indicate users' check-in state, such as a specific hour of a day (8:00), or a certain day of a week (Monday). By denoting the temporal state, the functions, $h(t)$ and $w(t)$, can output the hour of a day and the day of

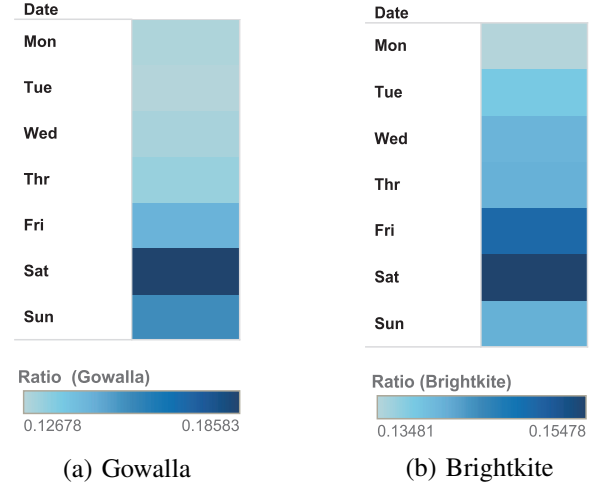


Fig. 2: Weekly Check-in Distribution of Gowalla and Brightkite Datasets.

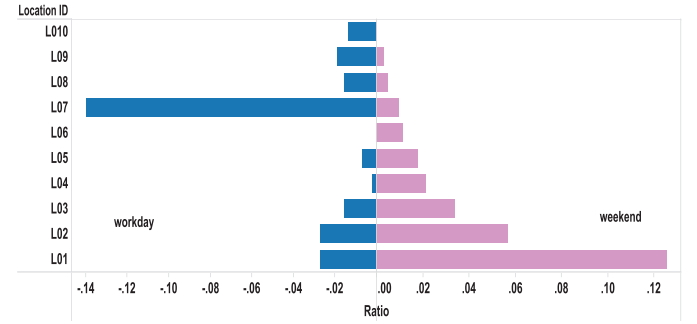


Fig. 3: The ratio of a random user check-in at his most frequent locations on workdays and weekend in Gowalla.

a week respectively, when input time t , e.g., $t = (2010-05-17\ 08:30:25)$, $h(t) = 8$.

a) **Daily Feature (DF)**: People tend to go to the same place at the similar time of a day as a routine activity. For example, a user goes to work at 9 am, therefore he usually checks in at the breakfast shop around 8 am. Thus we define DF as the check-in frequency of u at locations in $h(t) \in \{0, 1, \dots, 23\}$.

b) **Weekly Feature (WF)**: Previous works have discovered that human mobility exhibits strong similarity from Monday to Friday for working and strong diversity from Saturday to Sunday for amusing [16]. Fig. 2 plots the ratio of users' daily check-ins of a week from Gowalla and Brightkite, with "the deeper color, the more check-ins", proving the phenomenon of "check-in similarity during the workdays and check-in diversity on weekends". Meanwhile, in Fig. 3, we plot the distribution of a random user's historical check-ins at the locations with high check-in frequencies on workdays and weekends. The figure shows a clear preference difference between workdays and weekends for this user. Consequently, we define WF as weekly check-in frequency, with $w(t) = 0$ and $w(t) = 1$ indicating workdays and weekends, respectively.

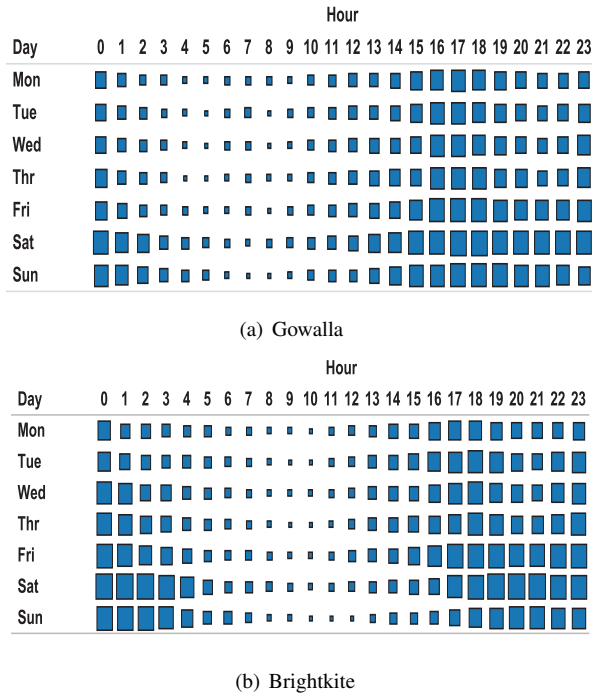


Fig. 4: Daily-Weekly Check-in Distribution of Gowalla and Brightkite Datasets

c) Daily-Weekly Feature (DWF): Human mobility exhibits that DF and WF are two fundamental temporal features in reflecting temporal cyclic patterns of users' mobile behavior [14]. In Fig. 4, each grid represents the number of check-in activities happened during any hour in each day of a week for Gowalla and Brightkite, with "the bigger grid, the more check-in activities". Fig. 4 shows both DF and WF features indicating mobile tendency on specific periods of time, e.g., a user usually visits the location near the company around 9 am on workdays, and generally stays close to home around 9 am on weekends. Then we define DWF as the check-in frequency in a particular hour of a particular day.

To investigate user's mobile behavior, each check-in record could be represented as a tuple $\langle u, w(t), h(t), l', l \rangle \in V_u$, indicating user $u \in U$ checks in at location $l \in L$ with latest location $l' \in L$ at time t , where V_u is a sequence of u 's check-ins in training set, $U = \{u_1, u_2, \dots, u_m\}$ is the set of users, and $L = \{l_1, l_2, \dots, l_n\}$ is the set of locations. Let $H_u = \{\langle u_i, w(t_q), h(t_q), v_{q-1}, v_q \rangle | \langle u, w(t), h(t), l', l \rangle \in V_u, u \in U, t \in T, \}$ be the set of u 's historical check-in actions, where v_q denotes the q -th check-in location, and v_{q-1} denotes the $(q-1)$ -th location.

III. MODELING LOCATION PREDICTION

Based on the analysis of mobile features, SIF, SMF, DF, WF and DWF, we first present a set of state-of-the-art approaches of location prediction. Then, considering user temporal feature and relationship between multi-features separately, we model various kinds of location prediction methods.

A. State-of-the-art Model

We introduce four traditional location prediction models in three aspects: spatial, temporal and spatio-temporal.

1) Spatial Model:

a) Most Frequent Check-in Model (MFC): Based on SIF, MFC [13] assigns the check-in probability at location l of a user u as the appearing probability of l in u 's historical check-ins.

b) Order-1 Markov Model (OMM): Based on SMF, OMM [19] assigns the searching result for the next location with the highest check-in probability according to the conditional probability of u 's check-ins at l based on l' 's as the prediction result.

2) Temporal Model:

a) Most Frequent Time Model (MFT): Based on DF, MFT [20] suggests the location with the maximum check-in probability at the time $h(t)$ of u as the prediction result for $h(t)$.

3) Spatio-Temporal Model:

a) HPY Prior Hourly Model (HPH): Considering SMF and DF, HPH [17] leverages GMM to predict next location at $h(t)$ through modeling users' mobile behavior in the context of the latest visiting location l' . The candidate location is the most likely to happen at $h(t)$ after l' with the maximum check-in probability.

B. Proposed Model

1) Single Feature Model: Previous works have modeled location prediction according to weekly periodicity of weeks' same day, the fine-granularity of temporal information leads to low prediction accuracy due to the sparsity of data. Thus, based on obvious WF and DWF features of users, we propose the following two single-feature models.

a) Most Frequent Week Model (MFW): Considering WF of user behavior, we propose MFW to capture temporal periodicity in workdays and weekends, defining u 's check-in probability at l on the day $w(t)$ as the following,

$$p_{MFW}(v_q | H_{u, w(t_q)=w}) = \frac{|\{v_k | v_k \in V, v_k = l, w\}|}{|\{v_k | v_k \in V, w\}|}, \quad (1)$$

where $H_{u, w(t_q)=w}$ represents the set of u 's check-ins on w in the history.

We select the locations with the maximum check-in probabilities as prediction results from all target locations for each test check-in. If there is only one candidate location produced in algorithm process, the prediction probability is 1 for this location. If the number of candidate locations is more than 1, multiple locations would be checked in with the same prediction probability, which is subject to the rule that the sum of the prediction probabilities is 1.

Example: A user u visited A, B, C, D and E in the history with probabilities 0.3, 0.1, 0.2, 0.3 and 0.1. Because of the highest probabilities in A and D , we predict u will check in at A and D with the prediction probability 0.5.

The predicted location set is as the following:

TABLE I: Available Features Map: *ApplicableModel*

<i>AvailableFeature</i>	I_{v_i}
SIF	1
SMF,SIF	1,2
DF,SIF	1,3
WF,SIF	1,4
DF,SMF,SIF	1,2,3
WF,SMF,SIF	1,2,4
DF,WF,DWF,SIF	1,3,4,5
DF,WF,DWF,SMF,SIF	1,2,3,4,5

$$L_P = \underset{l_j \in L}{\operatorname{argmax}} (p_{MFW}(l_j|H_{u,w})). \quad (2)$$

The prediction probabilities of candidate locations for u on w are defined as:

$$P_{MFW}(l_k|H_{u,w}, l_k \in L_P) = \frac{1}{|L_P|}, \quad (3)$$

let L_P be the set of prediction locations, $|L_P|$ be the cardinal number of L_P .

b) Most Frequent Daily-Weekly Model (MFDW): Based on DWF, we propose MFDW to predict a user's location, indicating u 's check-in probability at l at $h(t)$ on $w(t)$ as the following:

$$p_{MFDW}(v_q|H_{u,w(t_q)=w,h(t_q)=h}) = \frac{|\{v_k|v_k \in V, v_k = l, w, h\}|}{|\{v_k|v_k \in V, w, h\}|}, \quad (4)$$

where $H_{u,w(t_i)=w,h(t_i)=h}$ indicates the set of u 's check-ins at h on w in the history. And the prediction is similar to MFW,

$$L_P = \underset{l_j \in L}{\operatorname{argmax}} (p_{MFDW}(l_j|H_{u,w,h})), \quad (5)$$

$$P_{MFDW}(l_k|H_{u,w,h}, l_k \in L_P) = \frac{1}{|L_P|}. \quad (6)$$

2) Multi-Features Model: As modeling location prediction with single feature is unilateral, we pay more attention to how to utilize all features comprehensively.

The incomplete user-driven check-in property, for example, a user doesn't have historical check-ins in some hours of each day or other instances of lacking historical information, usually shows low prediction efficiency. Meanwhile, previous prediction models usually consider a small part of user features, resulting in low prediction accuracy. Since considering all five aspects features we have classified in Section II, and selecting the most powerful feature tendency can significantly improve the prediction effect, we propose a multi-features algorithm, **Powerful Feature Tendency Based Selection (PFTS)**.

For finding efficient features, we define users' available features map as *ApplicableModel*[*AvailableFeature*] = I_{v_i} , indicating the mapping from the features to the prediction models in Table I, where *AvailableFeature* is a set of features possessing historical information, and I_{v_i} is a set of acting models' labels with $\{1, 2, 3, 4, 5\}$ corresponding to

$\{MFC, OMM, MFT, MFW, MFDW\}$.

Example: For a check-in record, "1, 0, 10, 87, 223", we predict where the user "1" will visit at the time "10" on weekends after he last checked in at the location "87" with actually check-in at the location "223". If no historical check-ins at time "10" for user "1", user "1" doesn't have DF and DWF, meaning MFT and MFDW models don't work in this context. So *AvailableFeature* = $\{SIF, SMF, WF\}$, $I_{v_i} = \{1, 2, 4\}$, and then the acting models = $\{MFC, OMM, MFW\}$.

By the above analysis, we find that, based on the analysis of some feature, e.g., *SIF*, users visit a small portion of locations with high probabilities, and the remaining locations with low probabilities. However, based on others, e.g., *SMF* and *WF*, users visit all locations with similar probabilities. The non-uniform probability distribution indicates apparent mobile tendency, and results in a larger variance. Due to the distinction of feature-tendency, applying different features to predict locations leads to different performances of accuracy. Aiming to self-adaptively select the feature with the most powerful tendency, we use variance σ_n^2 ($n \in I_{v_i}$) of feature's probability distribution to measure tendency strength. In addition, our selection approach has strong universality for any feature set, which is convenient to introduce new features to improve the effectiveness of location prediction.

$$\sigma_n^2 = \frac{1}{N_{u,n}} \sum_{q=1}^{N_{u,n}} (p_n(v_q|H_{u,n}) - \mu_n)^2, \quad (7)$$

where μ_n is the mean of check-in probability as the following:

$$\mu_n = \frac{\sum_{q=1}^{N_{u,n}} p_n(v_q|H_{u,n})}{N_{u,n}}. \quad (8)$$

- $H_{u,1}$ is the set of u 's historical check-ins, correspondingly $N_{u,1}$ is the number of u 's total check-ins;
- $H_{u,2}$ indicates the set of u 's check-in records with the latest check-in at the location v_{q-1} in the history, accordingly $N_{u,2}$ is the number of these records;
- $H_{u,3}$ is the set of u 's historical check-ins at the time $h(t_q)$, and $N_{u,3}$ is the number of check-ins in this context;
- $H_{u,4}$ is same to $H_{u,w}$, and $N_{u,4}$ is the total number of u 's check-ins on w ;
- $H_{u,5}$ is same to $H_{u,w,h}$, as $N_{u,5}$ is the number of u 's check-ins at h on w .

Combining spatio-temporal information with feature tendency together, we provide personal prediction results for each record through selecting the feature with the most powerful tendency, indicated by the maximum variance:

$$M = \underset{n \in I_{v_i}, n \in SerNum}{\operatorname{argmax}} \sigma_n^2. \quad (9)$$

$$p_{PFTS}(l|H_{u,w,h,v_{q-1}}) = p_M(l|H_{u,w,h,v_{q-1}}). \quad (10)$$

Algorithm 1 *PFTS* Prediction Model

Input: a tuple t_1 of attributes
($\text{user id}, \text{day of week}, \text{hour of day}, \text{last location id}, \text{current location id}$)

Output: prediction results P_{PFTS}

```
1: Set  $features = \{DF, WF, DWF, SMF, SIF\}$ 
2: Set  $AvailableFeature = \text{empty}$ 
3: for each  $feature$  in  $features$  do
4:   if  $historicalInfo(feature)$  exists then
5:     add  $feature$  into  $AvailableFeature$ 
6:   end if
7: end for
8:  $I_{v_i} = ApplicableModel[AvailableFeature]$ 
9: for  $n$  in  $I_{v_i}$  do
10:   $[\mu_n, \sigma_n, P_n] \leftarrow f_n(t_1)$ 
11:   $M \leftarrow \arg \max_{n \in I_{v_i}} \sigma_n^2$ 
12: end for
13: for  $l_j$  in  $L$  do
14:   $L_P \leftarrow \arg \max_{l_j \in L} (p_M(l_j | H_{t_1}))$ 
15: end for
16: return  $P_{PFTS} = \frac{1}{|L_P|}$ 
```

We calculate prediction results as follows:

$$L_P = \underset{l_j \in L}{\operatorname{argmax}} (p_{PFTS}(l_j | H_{u,w,h,v_{q-1}})), \quad (11)$$

$$P_{PFTS}(l_k | H_{u,w,h,v_{q-1}}, l_k \in L_P) = \frac{1}{|L_P|}. \quad (12)$$

Based on the above analysis, a pseudocode description of *PFTS* model is given in **Algorithm 1**. First, initialize the feature sets, where *features* and *AvailableFeature* respectively represent the set of all features and the set of the available features (Lines 1-2). Then, preprocess data to search for available features with historical information (Lines 3-7). Next, record the mapping from *AvailableFeature* to *AvailableModel* by I_{v_i} (Line 8). Select the feature with the maximum variance of probability distribution, and utilize corresponding model M to predict location for check-in t_1 (Lines 9-12). Finally, provide candidate locations and return prediction probabilities (Lines 13-16)

IV. EXPERIMENTS

In this section, we evaluate the prediction performance of our proposed model. In particular, we perform the evaluation in two stages: (1) compare the prediction effects based on different combinations of mobile features; (2) estimate the proposed model in contrast to baseline models.

A. Datasets

In this paper, we utilize two publicly available LBS datasets, Gowalla and Brightkite, with the geographical distribution of check-ins in Fig. 5, to evaluate prediction models. We select users with at least 10 check-ins, then obtain 15,398 users from

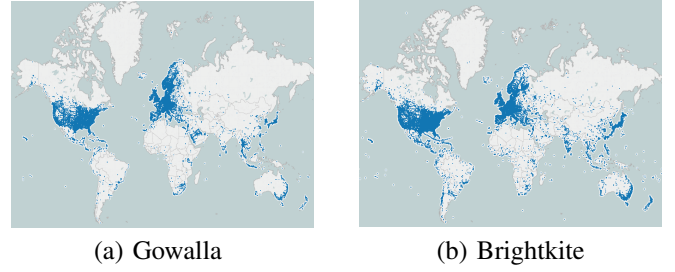


Fig. 5: The Check-in Geographical Distribution of Gowalla and Brightkite Datasets

TABLE II: Statistical Information of Two Datasets

	Gowalla	Brightkite
Duration	01/2011 - 09/2011	05/2009 - 10/2010
# of users	153,98	25,925
# of check-ins	3,969,935	4,576,761
# of unique locations	978,723	741,311

Gowalla and 25,925 users from Brightkite in Table II. For the sake of the diversity of datasets, we use the check-in data of September, 2011 as test set for Gowalla and randomly sample five check-ins for each user as test set for Brightkite.

Details of the Gowalla dataset are depicted as follows:

- The training set: containing 3,772,038 check-ins of 15,398 users at 974,992 locations.
- The testing set: including 197,897 check-ins of 15,398 users at 104,160 locations.
- The complete data set: covering 3,969,935 check-ins of 15,398 users at 978,723 locations.

Details of the Brightkite dataset are depicted as follows:

- The training set: covering 4,454,580 check-ins at 721,159 locations of 25,925 users.
- The testing set: including 129,625 check-ins at 72238 locations of 25,925 users.
- The complete data set: containing 4,576,761 check-ins of 25,925 users at 741,311 locations.

B. Evaluation Metrics

We predict locations at each check-in record $\langle u, w(t), h(t), l' \rangle$, with historical check-ins before the record time t as observed context. Based on prediction probability P_M , the accurate probability for a test check-in $c_i \in C = \{c_1, c_2, \dots, c_k\}$, where C is a sequence of users' check-ins in test set, is defined as:

$$P_{c_i} = \begin{cases} P_M(v_{c_i} = l_{c_i} | c_i) & (a) \\ 0 & (b) \end{cases}, \quad (13)$$

where Eq.(13) indicates accurate probability, Eq.(13).(a) for a correct prediction and Eq.(13).(b) for a wrong prediction.

The prediction accuracy (**PA**) is the average of P_{c_i} for all test check-ins,

$$PA(C) = \frac{\sum_{i=1}^{|C|} \{c_i | c \in C, P_{c_i}\}}{|C|}, \quad (14)$$

TABLE III: Location Prediction of PFTS Based on Different Combinations of Features

SIF	SMF	DF	WF	DWF	Gowalla	Brightkite
✓	✓	✓	✓	✓	0.1394	0.3984
✓	✓	✓	✓		0.1360	0.3976
✓	✓	✓		✓	0.1389	0.3966
✓	✓		✓	✓	0.1396	0.3991
✓	✓			✓	0.1388	0.3980
✓	✓	✓			0.1358	0.3965
✓	✓		✓		0.1225	0.3943
✓		✓			0.1263	0.3418
✓			✓		0.1057	0.3358
✓				✓	0.1291	0.3445
✓	✓				0.1187	0.3920
✓					0.1003	0.3294

TABLE IV: Comparison the Baselines Accuracy with PFTS

	Models	Gowalla	Brightkite
Spatial-based	MFC	10.03%	32.94%
	OMM	7.41%	33.47%
Temporal-based	MFT	12.01%	32.63%
	MFW	10.76%	32.62%
	MFDW	12.06%	32.52%
Spatio-temporal	HPH	8.07%	29.13%
	PFTS	13.96%	39.19%

where $|C|$ is the cardinal number of C .

C. Results and Discussions

Firstly, we investigate the prediction performance based on different combinations of mobile features in Gowalla and Brightkite. We can summarize several observations from Table III showing the detailed prediction effects:

- Only combining SIF with a temporal feature or SMF performs better in PA than using single feature, while worse than the fusion of all dimension features. Thus, the more dimension features into modeling human mobility, the more accurate predictions we have.
- DF and WF features both capture users' temporal preference. However, SIF+DWF outperforms SIF+DF and SIF+WF, suggesting these two features contain complementary information to describe users' mobile behavior.

Secondly, we evaluate the prediction performances of six baselines compared with **PFTS** in Table IV. We conclude the essential observations as below:

- Our proposed model, weighing both spatial and temporal information, outperforms the solely spatial-based or temporal-based models, which is actually significant in location prediction.
- The proposed model performs much better than spatial-based models in prediction-accuracy, outperforming MFC and OMM by 39.18% and 88.39% in Gowalla dataset, and 21.16% and 19.24% in Brightkite dataset. Moreover, compared with temporal-based models, **PFTS** shows extra 16.24%, 29.74% and 15.75% relative improvements over MFT, MFW and MFDW in Gowalla, and 22.31%, 18.7% and 22.72% higher in Brightkite. Obviously, spa-

tial and temporal information are fundamental for analyzing user behavior together.

- Although HPY considers both spatial (SMF) and temporal (DF) features, the PA of **PFTS** exceeds spatio-temporal model, HPY, by 72.99% and 37.01% in Gowalla and Brightkite respectively, due to the incorporation of both spatio-temporal features and feature tendency. The results indicate that **PFTS** uses spatio-temporal information more sufficiently. Moreover, the results demonstrate spatio-temporal context and feature-tendency are complementary driver in researching human mobility, which are helpful in improving the prediction accuracy.
- Consequently, **PFTS** consistently outperforms spatial-based, temporal-based and spatio-temporal-based models. Based on the ability of synthetically utilizing mobile information and automatically selecting the most powerful feature, our model can remarkably boost the prediction performance.

V. CONCLUSION AND FUTURE WORK

In this paper, we comprehensively investigate user's multidimensional mobile features, SIF, SMF, DF, WF and DWF, particularly temporal cyclic patterns for LBSNs. By integrating both spatio-temporal features and the selection of feature tendency, we have proposed an universal location prediction model called **PFTS**, which can self-adaptively select the feature with the most powerful tendency for various scenes. Under the guidance of the chosen feature, the corresponding method is utilized to predict location automatically. Based on the evaluation results, our model significantly outperforms existing approaches, which demonstrates the feature tendency can present user behavior perfectly and capture human mobility effectively.

In the future, we will exploit more users' mobile features and social influence modeling user behavior to achieve better performances of location prediction in LBSNs.

VI. ACKNOWLEDGEMENTS

This work is supported by the National High Technology Research and Development Program of China (863 Program) (No. 2013AA013603), the National Natural Science Foundation of China (No. 61471060) and the Funds for Creative Research Groups of China (No. 61421061).

REFERENCES

- [1] A. Monreale, F. Pinelli, R. Trasarti, and F. Giannotti, "Where next: a location predictor on trajectory pattern mining," *SIGKDD*, pp. 637–646, 2009.
- [2] L. Song, D. Kotz, R. Jain, and X. He, "Evaluating location predictors with extensive wi-fi mobility data," *INFOCOM*, vol. 2, pp. 1414–1424, 2004.
- [3] N. Yuan, Y. Wang, F. Zhang, X. Xie, and G. Sun, "Reconstructing individual mobility from smart card transactions: A space alignment approach," *ICDM*, pp. 877–886, 2013.
- [4] C. Song, Z. Qu, N. Blumm, and A. Barabási, "Limits of predictability in human mobility," *Science*, vol. 327, pp. 1018–1021, 2010.
- [5] P. Tian, A. V. Sanjay, K. Chiranjeevi, and S. M. Malik, "Intelligent advertising framework for digital signage," *KDD*, 2012.
- [6] E. Glaeser and M. Kahn, "Sprawl and urban growth," *Handbook of regional and urban economics*, 2004.

- [7] S. Eubank, H. Guclu, V. Anil, M. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang, "Modeling disease outbreaks in realistic urban social networks," *Nature*, 2004.
- [8] F. Boccardi, R. Heath, A. Lozano, T. Marzetta, and P. Popovski, "Five disruptive technology directions for 5g," *Communications Magazine, IEEE*, vol. 52, no. 2, pp. 74–80, 2013.
- [9] C. Chow and M. Mokbel, "Trajectory privacy in location-based services and data publication," *SIGKDD*, vol. 13, no. 1, pp. 19–29, 2011.
- [10] T. Kärkkäinen, M. Pitkänen, and J. Ott, "Enabling ad-hoc-style communication in public wlan hot-spots," *SIGMOBILE Mobile Computing and Communications Review*, vol. 17, no. 1, pp. 4–13, 2013.
- [11] S. Liu, Y. Liu, L. Ni, J. Fan, and M. Li, "Towards mobility-based clustering," *KDD*, pp. 919–928, 2010.
- [12] Y. Wang, N. J. Yuan, D. Lian, L. Xu, X. Xie, E. Chen, and Y. Rui, "Regularity and conformity: Location prediction using heterogeneous mobility data," *KDD*, pp. 1275–1284, 2015.
- [13] J. Chang and E. Sun, "Location3: how users share and respond to location-based data on social networking sites," *AAAI*, pp. 74–80, 2011.
- [14] M. Ye, K. Janowicz, C. Mülligann, and W. Lee, "What you are is when you are: The temporal dimension of feature type in location-based social networks," *SIGSPATIAL*, 2011.
- [15] E. Malmi, T. Do, and D. Perez, "Checking in or checked in: Comparing large-scale manual and automatic location disclosure patterns," *MUM*, 2012.
- [16] E. Cho, S. Myers, and J. Leskovec, "Friendship and mobility: user movement in location-based social networks," *KDD*, pp. 1082–1090, 2011.
- [17] H. Gao, J. Tang, and H. Liu, "Mobile location prediction in spatio-temporal context," *Nokia Mobile Data Challenge*, 2012.
- [18] H. Gao, J. Tang, X. Hu, and H. Liu, "Modeling temporal effects of human mobile behavior on location-based social networks," *CIKM*, 2013.
- [19] Y. Song, Z. Hu, X. Leng, H. Tang, K. Yang, and X. Ke, "Friendship influence on mobile behavior of location based social network users," *Communications and Networks*, vol. 17, no. 2, pp. 126–132, 2015.
- [20] G. Gidófalvi and F. Dong, "When and where next: Individual mobility prediction," *SIGSPATIAL MobiGIS*, pp. 57–64, 2012.



Tu Chen received B.E. degree in Communication Engineering from BUPT in 2013. He is currently a M.S. candidate in State Key Laboratory of Networking and Switching Technology, BUPT. His research interests include social networks, recommendation system, data mining.



Jing Zhang received B.E. degree in Communication Engineering from Xidian University in 2014. She is currently a M.S. candidate in State Key Laboratory of Networking and Switching Technology, BUPT. Her major interests are social networks, LBSN, HTML.



Zi Xing received B.E. degree in Communication Engineering from Communication University of China in 2014. She is currently a M.S. candidate in State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications (BUPT). Her major interests are social networks, user behavior, link prediction and mobile networks.



Hui Tian received her M.S. degree in Micro-Electronics and Ph.D. degree in Circuits and Systems both from Beijing University of Posts and Telecommunications, China in 1992 and 2003, respectively. Currently, she is a professor at BUPT. Her current research interests mainly include LTE/4G/5G system design, the MAC protocols, resource scheduling, cross-layer design and cooperative communication in such systems, heterogeneous wireless networks, ad hoc and sensor networks, and common radio resource.