Supporting Serendipitous Social Interaction Using Human Mobility Prediction

Zhiwen Yu, Hui Wang, Bin Guo, Tao Gu, and Tao Mei

***Abstract*—Leveraging the regularities of people's trajectories, mobility prediction can help forecast social interaction opportunities. In this paper, in order to facilitate real-world social interaction, we aim to predict "serendipitous" social interactions, which are defined as unplanned encounters and interaction opportunities and regarded as emerging social interactions. We collected GPS trajectory data from people' daily life on campus and use it as empirical mobility traces to generate decision trees and model trees to predict next venues, arrival times, and user encounter. Mobility regularities are mainly considered in these prediction models, and mobility contexts (e.g., time, location, and speed) act as decision nodes in the classification trees. Experimental results using collected GPS data showed that our system achieves 90% accuracy for predicting a user's next venue using a decision tree algorithm, with minute-level (around 5 min) prediction error for arrival time using the model tree algorithm. Two prototype applications were developed to support serendipitous social interaction on campus, and the feedback from a user study with 25 users demonstrated the usability of these two applications.**

***Index Terms*—GPS data, inference model, mobility prediction, serendipitous social interaction, user study.**

I. Introduction

With the rising popularity of mobile sensors and portable devices, "serendipitous" social interaction is emerging and becoming pervasive, where interaction is triggered when two devices (i.e., two users) are located closely in a mobile peer-to-peer environment [8], or wireless ATM networks [31]. Studies have been conducted to support unexpected interaction utilizing users' mobile wireless devices (e.g., mobile phone, vehicle PDA, and wearable devices). In these studies, human interaction is facilitated by means of peer-to-peer data sharing between matching devices [7] (i.e., users), and the "matching" is defined by similarity computed from user profiles (e.g., music, pictures, and tagged preferences) on the phone. However, these mechanisms merely provide services for people who have common interests and thus cannot enhance human social relations in the physical world.

The term "serendipity" was coined by the novelist Horace Walpole in the 18th century to describe unexpected and fortunate discoveries; "serendipity" was originally used to refer to making accidental discoveries when looking for one thing and finding another [2]. We, therefore, characterize the serendipitous social interaction as: unplanned, not affecting users' schedules, bringing convenience, creating positive emotion (e.g., happiness) in serving as an activity users, and users can choose to participate or not. For example, take two college students who are friends and who have not met for weeks. Both of them often visit the same library at the same time, but they do not recognize that. In such a scenario, if a mobile application helps them to capture and learn the serendipitous opportunities to meet and have a chat, the friendship between these two students will be enhanced.

To support such social interactions, we need first to discover the serendipitous interaction opportunities [21], similar to the communication channel of mobile intermediate nodes in an opportunistic network [3]. With the captured unplanned and transitory opportunities, serendipitous interaction can be used to make our lives easier. For example, if a user is aware that his roommate is passing by a grocery, he may ask him to buy something for him. However, this kind of serendipitous opportunity will disappear in a matter of seconds as she walks away. In this scenario, prediction, rather than instant behavior detection or notification, would be better.

Mobility prediction typically leverages human trajectory data (e.g., GPS data, check-in records, and intercell [33]). The majority of mobility prediction methods focus on predicting a user's future mobility status (e.g., where and when a user arrives at the next venue, and duration). In this paper, we predict users' future temporal and spatial contexts such as venue, arrival time, and user encounter. As user mobility status (e.g., location and time) may change in a very short period of time in the physical world, coarse-grained (e.g., hour-level temporal prediction error) and low accuracy (e.g., accuracy of next venue judgment) mobility prediction cannot satisfy the needs of practical applications. Aiming to achieve high accuracy and low error, we discover the strong spatial and temporal regularities in GPS trajectories and use supervised learning algorithms to train historical mobility instances (generated by crowd users).

Few applications have been proposed, especially any combining spatial and temporal prediction information. In this paper, we aim to facilitate serendipitous social interactions and the real participation activities in the physical world, which is different from simple information sharing mechanisms in existing traditional applications. Leveraging the overlap and regularity in collected GPS trajectories, we deploy supervised learning algorithms and achieve an accuracy of over 90% for predicting a user's next venue, and minute-level (i.e., an average of about 5 min) prediction error for arrival time.

Manuscript received November 7, 2014; revised February 24, 2015; accepted June 6, 2015. Date of publication July 16, 2015; date of current version November 12, 2015. This work was supported in part by the National Basic Research Program of China under Grant 2015CB352400; the National Natural Science Foundation of China under Grant 61222209, Grant 61373119, and Grant 61332005; the Specialized Research Fund for the Doctoral Program of Higher Education under Grant 20126102110043; and Microsoft. This paper was recommended by Associate Editor V. Fuccella.

Z. Yu, H. Wang, and B. Guo are with the School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: zhiwenyu@nwpu.edu.cn; wanghui@mail.nwpu.edu.cn; guob@nwpu.edu.cn).

T. Gu is with the School of Computer Science and IT, RMIT University, Melbourne, Vic. 3000, Australia (e-mail: tao.gu@rmit.edu.au).

T. Mei is with Microsoft Research, Beijing 100080, China (e-mail: tmei@microsoft.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/THMS.2015.2451515

By leveraging mobility prediction to forecast the occurrence of serendipitous interaction opportunities, we propose a three-layer framework to support serendipitous social interaction, develop two applications for use on a university campus, and conduct a survey about the application. The main contributions of this paper are summarized as follows.

1) We propose a system framework for supporting social interaction by means of facilitating users to participate in interaction activities in the physical world. Under this framework, mobility prediction is introduced to capture serendipitous interaction opportunities.
2) We leverage mobility prediction to discover serendipitous social interaction opportunities. Based on the spatiotemporal mobility regularities in users' trajectories, we first predict where and when a user will arrive, and then, we can determine if some users may soon encounter (i.e., have the same destination).
3) We develop the prototype of the proposed framework and build two applications based on the framework, **HelpBuy** and **EaTogether**, to support serendipitous social interaction on campus.

The rest of this paper is organized as follows. Section II reviews related work on human mobility prediction and social interaction enhancement. Section III introduces the framework and illustrates specific technical details of mobility prediction. Section IV presents two applications. We present experimental results and user study feedback in Section V. The discussion is presented in Section VI, and we conclude the paper in Section VII.

II. Related Work

Previous studies attempted to facilitate online social interactions in social networks. Under the premise that users who have the same interests are more willing to interact with each other (i.e., user homogeneity), studies try to help users find interaction with friends through personal preference similarity calculation in social networks, in which social interactions are facilitated in the form of learning [4], date [8], [12], and travel [9]. These kinds of applications only enhance social interaction within an online social network, rather than establish social connectivity in the physical world.

On the other hand, researchers focus on supporting face-to-face human interaction, in which social interactions are facilitated by leveraging serendipitous communication opportunities between mobile devices (i.e., users). *BlueFriend* [10] is an application that leverages Bluetooth to find friends among nearby users. *Bluedating* [12] provides localized dating services to help users find desired partners. Paradiso *et al.* [5] developed a badge system, which is equipped with wireless infrared and radio frequency networking, to facilitate social interaction between wearers. Lawrence *et al.* [7] developed three applications by exploiting the "co-presence" interactions (i.e., incidental interactions) between mobile devices. In the mobile peer-to-peer environment, as Yang *et al.* [9] proposed, information sharing and social interaction are facilitated by capturing serendipitous interaction opportunities. However, interaction opportunities are always discovered by device detection. Devices are the actual participant in these applications, rather than the humans themselves. In this paper, we predict serendipitous interaction opportunities leveraging mobility prediction using users' current mobility status and present mobile applications that facilitate people actively participating and interacting in the physical world.

Mobility prediction is the main supporting technology of our system. It aims to discover serendipitous interaction opportunities. Mobility prediction has been studied to perceive human future mobility status (e.g., next venue, arrival time, and duration) from different perspectives. Xiong *et al.* [1] took advantage of the similarity between people's trajectories and proposed collective behavioral patterns for improving prediction accuracy. Do and Gatica-Perez [13] adopted a factorized conditional model according to the extracted features, which can reduce the size of the parameter space as compared to conditional model. Cho *et al.* [14] discussed the contribution of a location-based social network and an individual user's periodic movement pattern in mobility prediction and developed a model that combines the periodic day-to-day movement patterns with the social movement effects coming from the friendship network. Noulas *et al.* [15] used check-in data in Foursquare and proposed a set of mobility prediction features to capture the factor that drives users' movement, and achieved around 90% prediction accuracy. Baumann *et al.* [16] analyzed the influence of temporal and spatial features in mobility prediction and predicted transitions. McInerney *et al.* [17] presented a Bayesian model of population mobility to tackle the data sparsity problem in mobility prediction. Song *et al.* [29] utilized the Order-2 Markov predictor with fallback and obtained a median accuracy of about 72% for users with long trace lengths. Song *et al.* [29] provided evidence that the prediction accuracy of an individual's next location had an upper bound of 93%. Lin *et al.* [30] reported on a study of GPS data-based mobility prediction accuracy and suggested that a predictability upper bound of 90% is able to support ubiquitous applications. In this paper, we not only predict an individual user's next venue, but the arrival time and multiple users' encounters based on users' current mobility features (CMFs). We achieved a high prediction accuracy of around 90%, which enables discovering opportunities to support serendipitous live interactions.

III. Serendipitous Social Interaction Supporting System

A. Framework

The framework consists of three layers, as shown in Fig. 1.

1) Data Layer: In the data layer, a GPS dataset of trajectories is collected from users by the GPS data logger. Each GPS point has several spatial and temporal attributes, such as date, time, latitude, longitude, and speed, and trajectories with thousands of GPS points are recorded. These mobility data reflect the regularity of users' daily visits (i.e., the GPS point sequences in some venues) and the movement (i.e., the GPS point sequences between two venues). After data denoising, historical mobility instances are extracted from the GPS dataset for mobility prediction.

2) Mobility Prediction Layer: Repeatable behaviors and patterns exist in the GPS trajectories. For example, at a campus crossroad during the morning, if a student is walking westward,

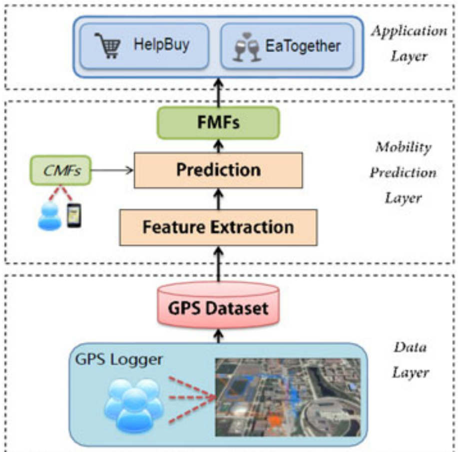


Fig. 1. System framework.

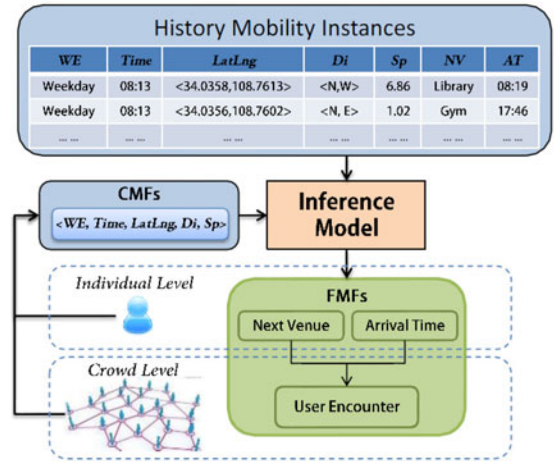


Fig. 2. Mobility prediction procedure.

then we predict that she will go to the gym; if she is walking toward the east, then we predict that she will arrive at the library. Moreover, people living in the same environment (e.g., students in a dormitory) may have the same regularity, and thus, users' mobility trajectories can be utilized as historical instances to predict the next venue and arrival time.

Human GPS trajectory contains many temporal (i.e., morning) and spatial (i.e., at a crossroad) context features. It is important to select the appropriate features to achieve high prediction accuracy and low complexity. As trajectories have regularity, we apply learning algorithms to discover patterns. Then, after perceiving individual users' CMFs (e.g., time, latitude, and longitude), the inference models can predict users' future mobility features (FMFs) (e.g., next venue and arrival time). On the individual level, we can predict a user's future status. Then, on the crowd level, we can predict multiple users' occasional encounters at a next venue (e.g., two users will meet at a restaurant in ten minutes), which is similar to predicting data source location in [11].

3) Application Layer: In this layer, we design and develop applications to facilitate serendipitous interactions based on the prediction results of user's FMFs from the mobility prediction layer. For example, we can develop more efficient participatory sensing [18] applications with perceiving users' future positions. Furthermore, we can discover unexpected opportunities with user encounter prediction results. In this paper, we implement two applications that facilitate social interactions in specific scenarios on a campus.

B. Mobility Prediction

We observe repeatable behavior and patterns in the GPS trajectories. For instance, a person is going to the same next venue when she/he is at the same position at the same time on different days (e.g., at noon, students at the same place are likely to go to restaurant; in the five minutes before class, students on different paths are all moving toward the classrooms). These real phenomena imply that students' trajectories contain strong spatial and temporal regularities. We explore these regularities for mobility prediction and apply supervised learning algorithms to train the inference model.

TABLE I
MOBILITY FEATURES

Features	Description	Example
@WE	weekday or weekend	weekday/ weekend
@Time	current time	18:06:08
@LatLng	latitude and longitude	<34.0356,108. 7600>
@Di (direction)	N(north)/S(south), E(east)/W(west)	<N, E>
@Sp (speed)	discretized into 4 levels, separated by 2, 5 and 10 (km/h)	Stroll(0–2)/Walk(2–5)/ Scurry(5–10)/Trot(> 10)
@NV	next venue	Library
@AT	arrival time of next venue	18:14:34
@UE	who, where and when will meet	< Tom,Restaurant, 11:56:32 >

In this paper, we predict users' FMFs at two granularities: individual and crowd. At the individual level, we predict where a user is going to (i.e., next venue) and when the user will arrive there (i.e., arrival time). At the crowd level, we predict the intersection of multiple users' trajectories (i.e., user encounter) based on the individual user mobility prediction.

Fig. 2 shows the detailed mobility prediction procedure. First, historical mobility instances are extracted from the historical GPS dataset to form the training dataset. Second, since supporting serendipitous social interaction requires high prediction accuracy both spatially and temporally, we adopt learning algorithms to train using the historical dataset. The prediction task contains discrete (next venue) and continuous (arrival time) output; therefore, we use different learning algorithms to train inference models to predict each FMF. According to [16], more CMFs are not necessary for higher prediction accuracy. We find that we need to select the suitable CMF set to achieve high accuracy and low complexity. Experiments are conducted by selecting different CMF sets from Table I and comparing their performance. Finally, after predicting results of the user's FMFs (i.e., next venue and arrival time) with the inference models, we can predict user encounter at the next venue.

1) Mobility Features: Taking every GPS point in the trajectories as a mobility instance, we extract fine-grained attributes (e.g., latitude, longitude, time, direction, and speed of a GPS point) as the mobility instance's features. Leveraging the overlap and regularity in these trajectories, the mobility instances' features are highly cohesive and exhibit low coupling.

Essentially, each mobility point is a GPS point recorded on the path between two venues with a timestamp and GPS coordinate. However, in this paper, a mobility point in fact has eight features (see Table I). The former five features are all CMFs, which are originally recorded by GPS sensors and the latter three features are all FMFs, @*NV*, @*AT*, @*UE*, which are acquired. Specifically, the historical mobility points' FMF values are assigned using feature extraction steps to form training instances, while the mobile phone users' FMFs values are to be predicted through inference models in practical applications. The CMFs are mutually independent.

None of the mobility features are related to personal information as limited by the data scale of individual users. Inspired by the *collective behavior pattern* [1] and the population modeling mechanism [11], we are able to achieve high prediction accuracy to meet the demands of application scenarios.

2) Predicting Next Venue: People's daily life generates almost the same life track every day, and their GPS trajectories exhibit high spatial and temporal overlapping and regularities. Specifically, a user's next venue is predominantly determined by his/her current spatial and temporal context (e.g., time and location), and people are likely to make the same decision when they are in the same situation (e.g., same location and same time). Therefore, to evaluate between users' mobility features and next venue, we choose Decision Tree [25], [26], Random Forest [22], KNN [22], and BayesNet [27] as candidate prediction models. These four supervised classification algorithms are implemented in WEKA [23]. By deploying these algorithms on historical mobility instances, we obtain inference models (see Fig. 2). These models take a user's mobility contexts (i.e., CMFs, e.g., time, latitude, longitude) as input conditions and determine the venue as output. We use tenfold cross-validation correct percentage to measure accuracy of predicting the next venue.

3) Predicting Arrival Time: Predicting arrival time accurately is crucial. For example, predicting users' future location and arrival time provides location based services to have a sufficient start-up time. Similar to the next venue prediction procedure, arrival time is also related to the user's mobility context; therefore, we need supervised classification algorithms to determine the arrival time. As the output of arrival time is a continuous value, we employ linear regression [22] and model tree [17], [24] in WEKA [23], to predict the arrival time. We use tenfold cross-validation to compare their performance in terms of arrival time prediction error.

4) Predicting User Encounter: If the users' mobility features, such as *next venue* and *arrival time*, can be predicted with high accuracy, then we predict people's encounters. We define of encounter as follows:

$$\text{Encounter}(s1, s2) \Leftrightarrow NV_1 = NV_2 \& \& |AT_1 - AT_2| < To. \quad (1)$$

In (1), $s1$ and $s2$ are two users, $NV1$ and $NV2$ are the predicted next venues of the two users, $AT1$ and $AT2$ are the predicted arrival time, and To is the preset time threshold that the former arriving user would wait for the later one. Combined with friend relationship information (i.e., friendship network as shown in Fig. 2), the encounter prediction results can be utilized in applications to support serendipitous social interactions between friends, unplanned in advance.

Overall, serendipitous social interaction opportunities (e.g., people encounters) are discovered from mobility prediction results and utilized in specific applications in the upper layer to support and enhance human interaction.

IV. Applications

We develop the prototype of the proposed framework. In this section, we present two prototype applications built on the framework to support serendipitous interactions on a campus. **HelpBuy** is a mutual application whereby a user requests others to buy what she/he needs, while others are incidentally on their way nearby a point of sale. **EaTogether** provides reminder services for users to find unplanned chance to eat together with their friends.

A. HelpBuy

HelpBuy is an application taking advantage of the next venue prediction. A user can request another to buy something along their way.

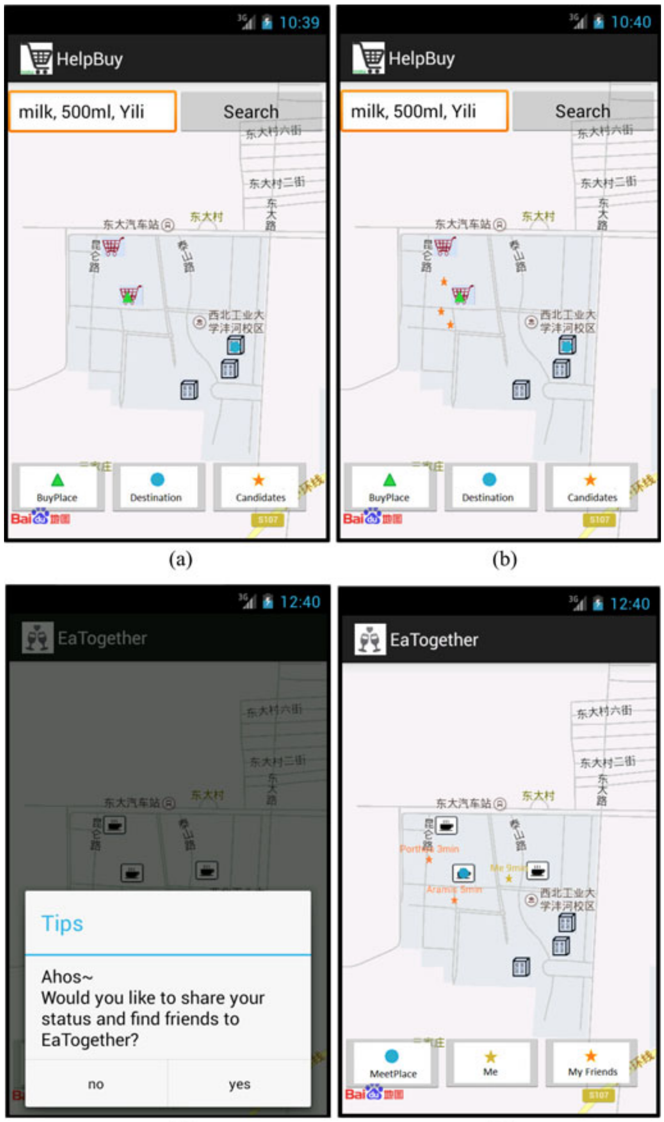
Usage scenario: One morning, Henry is working hard in the school laboratory. He wants a cup of milk, but he prefers not to leave the lab to go to the supermarket, which is far away. First, as shown in Fig. 3(a), Henry points out where to buy the milk on the map (*BuyPlace*,▲) and where the milk should be delivered to (*Destination*,◯ , i.e., Henry's current position, the lab building). He also needs to provide a description of what to buy (i.e., milk, 500 mL) and click the "Search" button. Then, as shown in Fig. 3(b), the users who are willing to help buy (i.e., volunteers, $\star$, near the *BuyPlace* and predicted to arrive at the *Destination*) appear on the map. Finally, one of the volunteers accepts Henry's request and fulfills the **HelpBuy** task.

In this application, people's willingness to share their real-time mobility status is critical. In order to encourage people (acting as the role of volunteer) to help others (acting as the role of clients), an incentive mechanism is needed. The client offers a reward when he/she publishes a task, and the volunteer earns reward points by fulfilling tasks.

B. EaTogether

EaTogether is an application that predicts when friends can have encounters at restaurants and recommends them to have a meal together.

Usage scenario: Three students, Athos, Aramis, and Porthos, are friends. They study on the same campus but they have not



(a) (b) (c) (d)

Fig. 3. Two applications. (a) HelpBuy (Before search). (b) HelpBuy (After search). (c) EaTogether (Tips). (d) EaTogether (Encounter notice).

met together for a long time. The system discovers if there is a chance for them to meet and have a chat with each other, for example, occasionally meeting at a campus restaurant and having lunch together. In the past, they had their lunches at the same restaurant at the same time, but not aware. Now, with **EaTogether**, they can capture every such opportunity to meet and have a meal.

One day at noon, **EaTogether** predicts that Athos is on the way to the restaurant and, then, asks Athos whether he wants to share his mobility status and try to discover unplanned chances for having lunch together with his friends as shown in Fig. 3(c). With Athos' permission, the backend server predicts that his friends, Aramis and Porthos, are heading toward the same destination, and then refreshes the screen of Athos's phone with an encounter notice, as shown in Fig. 3(d), in which *MeetPlace*, **, means the venue where they will meet (i.e., their same *next venue*), ⋆ denotes the user himself (i.e., Athos), while ⋆ denotes

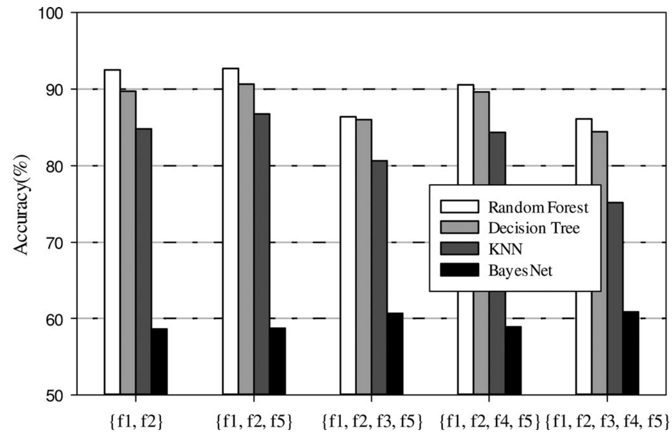


Fig. 4. Inference model performance comparison. (f1: Time, f2: LatLng, f3: Direction, f4: Speed, f5: WE).

his friends. In this application, we also show when the user and his friends will arrive at the venue.

Apart from *next venue* and *arrival time* prediction results, the users' relationship information is needed in this application. We can collect users' friendship information from online social networks, e.g., Facebook. By facilitating friends' unplanned interaction in this way, **EaTogether** is able to enhance, or strengthen, their friendship.

V. Experimental Results

A. Data Collection

We collected GPS trajectory data generated by volunteer students in a 1000 $\times$ 1000 m^2 campus. We selected 17 popular venues—classified into four categories: restaurant, dormitory, workplace, and sports—as the next venue candidates. Each of the student volunteers carried a portable GPS logging device all day, and the device recorded the student's daily trajectories.

The dataset includes trajectories from 156 student volunteers. The trajectories last 1200 hours, and there are about 2.8 million GPS points in the GPS dataset. These GPS points are converted to 64 482 mobility instances after preprocessing, and these historical instances, with CMF features and FMF features, act as prior empirical instances to train inference models. Utilizing these models, we predict FMFs based on a user's CMFs (e.g., time, latitude, and longitude).

B. Results of Next Venue Prediction

We use tenfold cross-validation on 64 482 mobility instances to test the performance of different prediction models. Fig. 4 demonstrates the performance of four prediction models under five different CMF sets. Since the campus students' lifestyle is relatively fixed, the collected GPS trajectories contain strong temporal and spatial regularities. Therefore, we achieved good prediction performance after adopting supervised classification algorithms. From this, we see that, in general, random forest, decision tree, and KNN yield almost the same excellent performance, while BayesNet performs poorly since the features

TABLE II
TRAINING DECISION TREE ON DIFFERENT CMF SETS

Feature sets	Leaves Number	Size of the tree	Correct percentage
{f1}	497	993	35.46%
{f2}	922	1843	48.59%
{f1, f2}	4621	9241	89.73%
{f1, f2, f3}	5405	10 547	84.82%
{f1, f2, f4}	5198	9849	88.64%
{f1, f2, f5}	4388	8775	**90.62%**
{f1, f2, f3, f5}	5337	10 415	85.99%
{f1, f2, f4, f5}	4940	9397	89.64%
{f1, f2, f3, f4, f5}	5775	10 730	84.47%

(f1: Time, f2: LatLng, f3: Direction, f4: Speed, f5: WE).

are mutually independent. Considering the complexity of the inference model, we finally select decision trees to predict next venue.

Using decision trees, for feature selection, to achieve high prediction accuracy and low complexity, we need to compare the performance of different feature sets. Table II illustrates the training results of different CMF sets. The leftmost column shows the feature set, the middle two columns are from the decision tree, and the rightmost column indicates the tenfold cross-validation correct percentage.

From Table II, Time and Latlng are pivotal features in predicting the next venue, confirming the importance of temporal and spatial context in mobility prediction research. We see that the Direction and Speed features show negative effects; this may be due to GPS signal noise. Since college students' daily schedules are different on weekdays and weekends, the WE feature contributes to a 1% improvement accuracy. Considering the complexity of the decision tree and prediction accuracy, the feature set {f1, f2, f5} is selected as the CMFs to predict next venue.

We randomly selected 2000 mobility points as test instances to measure the prediction accuracy of the inference model. The overall accuracy was 92.1%. For the different venues, the prediction accuracies were 90.6% (restaurant), 88.1% (dormitory), 94.7% (workplace), and 91.5% (sports). In general, the next venue prediction accuracy is higher than 90%. Due to relatively dense buildings in the dormitory area, GPS signal noise makes it difficult to predict to which dormitory building the user will go. Note that the decision tree model needs 6 min for the 2000 next venue prediction task.

Fig. 5 shows the prediction accuracy of next venue with different numbers of known mobility points. If we know more about users' current mobility status (e.g., continuous mobility points collected on user's mobile phone), we can predict next venue more accurately.

Nguyen *et al.* [28] also adopted classification algorithms, decision tree, and KNN by using the API of WEKA, to predict the next location. Their reported training accuracies, respectively, 69.0% (decision tree, C4.5/J48) and 54.1% (KNN, K is empirically set to 5), was lower than ours. The reason is that our GPS data have better regularity and contain a set of fine-grained spatial and temporal mobility features. Our results also comply with the suggested predictability upper bound, i.e., 90% [30].

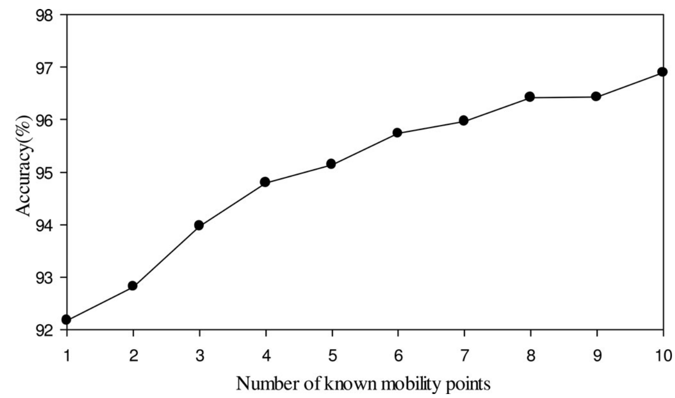


Fig. 5. Next venue prediction accuracy with different number of known mobility points.

TABLE III
REGRESSION PERFORMANCE COMPARISON

Feature sets	Correlation coefficient		Mean absolute error (hour)	
	Model tree	Linear regression	Model tree	Linear regression
{f1}	0.9983	0.9982	0.1459	0.154
{f2}	0.3823	0.1492	3.4134	3.8009
{f1, f2}	0.9994	0.9982	0.0992	0.1519
{f1, f2, f5}	0.9994	0.9982	0.0991	0.1518
{f1, f2, f3, f5}	0.9994	0.9983	0.099	0.1475
{f1, f2, f3, f4, f5}	0.9994	0.9983	0.0986	0.1476

(f1: Time, f2: LatLng, f3: Direction, f4: Speed, f5: WE).

C. Results of Arrival Time Prediction

In the arrival time prediction, we need a model to produce continuous values. Thus, we adopted two supervised learning algorithms as candidates: model tree and linear regression. We compare their performance by using predicted error of arrival time (AT):

$$\mathrm{error}_{\mathrm{AT}} = \mathrm{Predicted}_{\mathrm{AT}} - \mathrm{Real}_{\mathrm{AT}}. \quad (2)$$

Model tree performs better with a higher correlation coefficient and smaller mean absolute error (see Table III). Thus, we chose model tree as the training model in the arrival time prediction. As for feature selection, we selected {f1, f2, f3, f4, f5} as the CMFs to achieve the minimum mean error. The least mean absolute error is 0.0986 h (i.e., about 6 min) in this case.

We randomly selected 4000 mobility points as the test dataset to evaluate the arrival time prediction error and demonstrated the prediction error distribution as a CDF curve in Fig. 6. The errors are less than 10 min in most of the cases (more than 85%).

D. Results of User Encounter Prediction

Using (1), we selected 544 encounter instances to evaluate the user encounter prediction accuracy. In these instances, two users have an encounter after one waits roughly 2 min (i.e., To is set as 2 min) for the other. In our prediction validation, we successfully predicted 92.3% (502 of 544) of the instances.

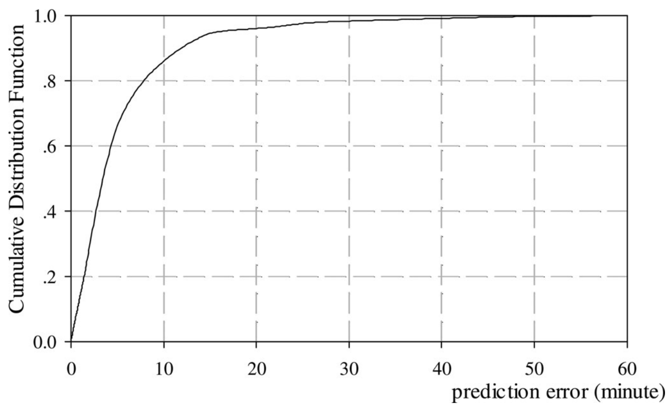


Fig. 6. CDF curve of arrival time prediction absolute error.

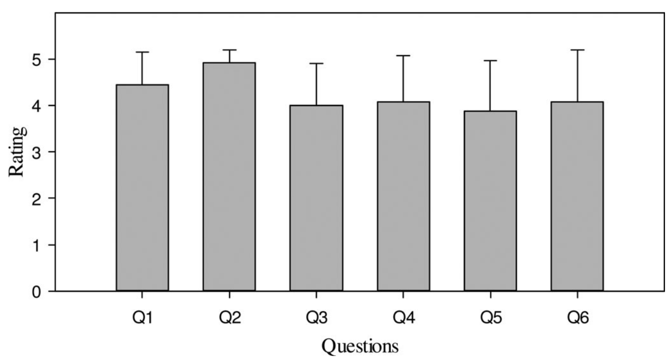


Fig. 7. User study results (the mean value and the standard error bars).

E. User Study

Twenty-five students (20 males and five females, majoring in a variety of subjects such as architecture and mathematics) were invited to use our applications for three days and then answer our questions. We collected the following feedback.

Q1: When people were caught in a similar situation as Henry, the **HelpBuy** application is useful.
Q2: If I encounter my friends occasionally at a campus restaurant, I like to have a meal together with them.
Q3: I would like to participate in **HelpBuy** activities, in which I can get others to help and I am willing to help others.
Q4: **EaTogether** helps to capture unplanned opportunities to meet friends, and I would like to use it again.
Q5: The applications perform well, and I was satisfied with the performance (prediction accuracy and prediction error) and the operating experience.
Q6: The applications strengthen the relationship between friends, and even strangers.

All the questions were answered using the following scale: 5 = strongly agree, 4 = agree, 3 = neutral, 2 = disagree, 1 = strongly disagree.

Fig. 7 illustrates the user study results. In general, the two applications, **HelpBuy** and **EaTogether**, are of practical use in people's lives (respectively illustrated by Q1 and Q2). The users' ratings of Q3 and Q4 indicate that people are willing to use the applications, but may have concerns. The results from Q5 and Q6 demonstrate that the applications may be useful in maintaining a closer relationship between people leveraging the serendipitous social interaction.

Some participants expressed their concerns about user privacy, and some believed that the user interface should be more design-friendly. Furthermore, some participants suggested more application scenarios and features to enrich our work, such as help returning books to the library, and to broaden application scenario to city-wide scale.

VI. Discussion

A. Offline Social Interaction

Traditional social interaction-supporting applications mainly focus on building an information exchange platform, without users getting involved. Offline social interactions not only can facilitate better relationship-building, but also help make full use of human resources by means of detecting to assistant opportunities (e.g., **HelpBuy** scenarios).

B. System Portability

To support upper-layer applications, we need high prediction accuracy. In our experiments, the strong spatiotemporal regularity in students' life trajectories may lead to acceptable prediction performance. If applied in a scenario where there is no such regularity, the prediction algorithms need to be further evaluated. With the development of localization technology, we may more precisely perceive users' locations and destinations; then, we can design and develop more advanced applications.

C. Privacy

Privacy is an inevitable issue in People-Nearby application [18]. In **HelpBuy**, we need to protect the information about the client and his or her goods. In **EaTogether**, a user may just want to disclose his/her mobility information to specific people.

VII. Conclusion

In this paper, we have proposed WeMeet framework that aims at supporting serendipitous social interaction leveraging human mobility prediction on campus. Using GPS trajectory traces, we designed a mobility prediction algorithm to predict people's mobility. Two prototype applications were implemented to support social interactions capitalizing on unplanned opportunities. **HelpBuy** is an interactive application to request "volunteers" to help buy something. **EaTogether** is another applied service to alert friends to eat together when occasional interaction opportunities appears.

The data were collected in a very specific scenario, i.e., students moving in a campus. In addition, we achieved a good prediction accuracy benefiting from students' repeatable trajectories. In the future, we will attempt to deploy the framework on other situations and other datasets with more occasional positive situations.

REFERENCES

[1] H. Xiong, D. Zhang, D. Zhang, and V. Gauthier, "Predicting mobile phone user locations by exploiting collective behavioral patterns," in *Proc. 9th Int. Conf. Ubiquitous Intell. Comput.*, 2012, pp. 164–171.

[2] I. Buchem, "Serendipitous learning: recognizing and fostering the potential of microblogging," *Formare Open J. per la formazione rete*, vol. 11, no. 74, pp. 7–16, 2011.

[3] C. M. Huang, K. C. Lan, and C. Z. Tsai, "A survey of opportunistic networks," in *Proc. 22nd Int. Conf. Adv. Inf. Netw. Appl.*, 2008, pp. 1672–1677.

[4] B. Thoms, "A dynamic social feedback system to support learning and social interaction in higher education," *IEEE Trans. Learning Technol.*, vol. 4, no. 4, pp. 340–352, Oct.–Dec. 2011.

[5] J. A. Paradiso, J. Gips, M. Laibowitz, S. Sadi, D. Merrill, R. Aylward, P. Maes, and A. Pentland, "Identifying and facilitating social interaction with a wearable wireless sensor network," *Pers. Ubiquitous Comput.*, vol. 14, no. 2, pp. 137–152, 2010.

[6] F. S. Tsai, W. Han, J. Xu, and H. C. Chua, "Design and development of a mobile peer-to-peer social networking application," *Expert Syst. Appl.*, vol. 36, no. 8, pp. 11077–11087, 2009.

[7] J. Lawrence, T. R. Payne, and R. V. Kripalani, "Exploiting incidental interactions between mobile devices," *MIRW*, vol. 56, pp. 56–59, 2006.

[8] S. D. Nagowah, "Aiding social interaction via a mobile peer to peer network," in *Proc. 4th Int. Conf. Digital Soc.*, 2010, pp. 130–135.

[9] W. S. Yang, S. Y. Hwang, and Y. W. Shih, "Facilitating information sharing and social interaction in mobile peer-to-peer environment," in *Proc. Pervasive Comput. Commun. Workshops*, 2012, pp. 673–678.

[10] P. Tamarit, C. T. Calafate, J. C. Cano, and P. Manzoni, "BlueFriend: Using bluetooth technology for mobile social networking," in *Proc. 6th Annu. Int. Conf. Mobile Ubiquitous Syst.: Comput., Netw. Services*, 2009, pp. 1–2.

[11] B. Resch, F. Calabrese, A. Biderman, and C. Ratti, "An approach towards real-time data exchange platform system architecture," in *Proc. 6th Annu. IEEE Int. Conf. Pervasive Comput. Commun.*, 2008, pp. 153–159.

[12] R. Beale, "Supporting social interaction with smart phones," *IEEE Pervasive Comput.*, vol. 4, no. 2, pp. 35–41, Jan.–Mar. 2005.

[13] T. M. T. Do and D. Gatica-Perez, "Contextual conditional models for smartphone-based human mobility prediction," in *Proc. ACM Conf. Ubiquitous Comput.*, 2012, pp. 163–172.

[14] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: User movement in location-based social networks," in *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2011, pp. 1082–1090.

[15] A. Noulas, S. Scellato, N. Lathia, and C. Mascolo, "Mining user mobility features for next place prediction in location-based services," in *Proc. IEEE 12th Int. Conf. Data Mining*, 2012, pp. 1038–1043.

[16] P. Baumann, W. Kleiminger, and S. Santini, "The influence of temporal and spatial features on the performance of next-place prediction algorithms," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2013, pp. 449–458.

[17] J. McInerney, J. Zheng, A. Rogers, and N. R. Jennings, "Modelling heterogeneous location habits in human populations for location prediction under data sparsity," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2013, pp. 469–478.

[18] E. Toch and I. Levi, "Locality and privacy in people-nearby applications," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2013, pp. 539–548.

[19] J. A. Burke, D. Estrin, M. Hansen, A. Parker. N. Ramanathan, S. Reddy, and M. B. Srivastava, "Participatory sensing," in *Proc. Int. Workshop World-Sensor-Web*, 2006, pp. 117–134.

[20] E. Paulos and E. Goodman, "The familiar stranger: Anxiety, comfort, and play in public places," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2004, pp. 223–230.

[21] C. Brown, C. Efstratiou, I. Leontiadis, D. Quercia, and C. Mascolo, "Tracking serendipitous interactions: How individual cultures shape the office," in *Proc. 17th ACM Conf. Comput. Supported Cooperative Work Social Comput.*, 2014, pp. 1072–1081.

[22] I. Witten, *Data Mining—Practical Machine Learning Tools and Techniques*, 2nd ed., San Francisco, CA, USA: Morgan Kaufmann, 2005.

[23] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. Witten, "The WEKA data mining software: An update," *ACM SIGKDD Explorations Newslett.*, vol. 11, no. 1, pp. 10–18, 2009.

[24] J. Quinlan, "Learning with continuous classes," in *Proc. Australian Joint Conf. Artif. Intell.*, 1992, pp. 343–348.

[25] J. Quinlan, *C4.5: Programs for Machine Learning*. San Mateo, CA, USA: Morgan Kaufmann, 1993.

[26] J. Quinlan, "Introduction of decision tree," *Mach. Learning*, vol. 1, no. 1, pp. 81–106, 1986.

[27] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," *Mach. Learning*, vol. 29, pp. 131–163, 1997.

[28] L. Nguyen, H. T. Cheng, P. Wu, S. Buthpitiya, and Y. Zhang, "PnLUM: System for prediction of next location for users with mobility," in *Proc. Nokia Mobile Data Challenge Workshop*, vol. 2, 2012.

[29] C. Song, Z. Qu, N. Blumm, and A.-L. Barabsi, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010.

[30] M. Lin, W. J. Hsu, and Z. Q. Lee, "Predictability of individuals' mobility with high-resolution positioning data," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2012, pp. 381–390.

[31] T. Liu, B. Paramvir, and C. Imrich, "Mobility modeling, location tracking, and trajectory prediction in wireless ATM networks," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 6, pp. 922–936, Aug. 1998.

[32] L. Song, D. Kotz, R. Jain, and X. He, "Evaluating location predictors with extensive Wi-Fi mobility data," in *Proc. 23rd Annu. Joint Conf. IEEE Comput. Commun. Soc.*, 2004, pp. 1414–1424.

[33] G. Yavaş, D. Katsaros, Ö. Ulusoy, and Y. Manolopoulos, "A data mining approach for location prediction in mobile environments," *Data Knowl. Eng.*, vol. 54, no. 2, pp. 121–146, 2005.