UNIVERSITY OF WEST LONDON
The Career University

UWL REPOSITORY

repository.uwl.ac.uk

Personalized location prediction for group travellers from spatial-temporal trajectories

Naserian, Elahe, Wang, Xinheng ORCID logoORCID: https://orcid.org/0000-0001-8771-8901, Dahal, Keshav, Wang, Zhi and Wang, Zaijian (2018) Personalized location prediction for group travellers from spatial-temporal trajectories. Future Generation Computer Systems, 83. pp. 278-292. ISSN 0167-739X

http://dx.doi.org/10.1016/j.future.2018.01.024

This is the Accepted Version of the final output.

UWL repository link: https://repository.uwl.ac.uk/id/eprint/4340/

Alternative formats: If you require this document in an alternative format, please contact: open.research@uwl.ac.uk

Copyright: Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy: If you believe that this document breaches copyright, please contact us at open.research@uwl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Rights Retention Statement:

Personalized Location Prediction for Group Travellers from Spatial-Temporal Trajectories

Elahe Naserian*, Xinheng Wang†, Keshav Dahal*, Zhi Wang‡, and Zaijian Wang§
*School of Engineering and Computing
University of the West of Scotland, Paisley, Scotland, UK
Email: {elahe.naserian, keshav.dahal}@uws.ac.uk
†School of Computing and Engineering
University of West London, London, UK
Email: {xinheng.wang}@uwl.ac.uk
‡Zhejiang University, Hangzhou, China
Email: wangzhi@iipc.zju.edu.cn
§Anhui Normal University, Wuhu, China
Email: wangzaijian@ustc.edu

***Abstract*—In recent years, research on location predictions by mining trajectories of users has attracted a lot of attentions. Existing studies on this topic mostly focus on individual movements, considering the trajectories as solo movements. However, a user usually does not visit locations just for the personal interest. The preference of a travel group has significant impacts on the places they have visited. In this paper, we propose a novel personalized location prediction approach which further takes into account users' travel group type. To achieve this goal, we propose a new group pattern discovery approach to extract the travel groups from spatial-temporal trajectories of users. Type of the discovered groups, then, are identified through utilizing the profile information of the group members. The core idea underlying our proposal is the discovery of significant movement patterns of users to capture frequent movements by considering the group types. Finally, the problem of location prediction is formulated as an estimation of the probability of a given user visiting a given location based on his/her current movement and his/her group type. To the best of our knowledge, this is the first work on location prediction based on trajectory pattern mining that investigates the influence of travel group type. By means of a comprehensive evaluation using various datasets, we show that our proposed location prediction framework achieves significantly higher performance than previous location prediction methods.**

***Index Terms*—Personalized location prediction, group pattern discovery ,trajectory mining, frequent movement patterns.**

I. Introduction

With the rapid development of mobile devices and location acquisition technologies, an enormous amount of trajectory data recording the movement of people is available. These overwhelming amounts of data is tremendously useful for the rapidly growing location-based applications market. Due to various requirements of these applications, e.g., system efficiency and marketing efficacy, accurately predicting the next location to which a user may move is essential. The location prediction technique identifies the next location that is most likely to be visited by the user, according to a set of application-dependent locations or pre-determined locations. By knowing the next movement of users, resources can be efficiently allocated to the most possible location, rather than the blind resource allocation. Efficient resource allocation to mobile users would lead to higher resource utilization and lower latency in accessing the resources. In addition, predicting the subsequent location can provide the insights for many existing pervasive applications, such as targeted advertising and services recommendation [1].

The problem of predicting the next location where a user will move has received many research interests in recent years. As the location prediction process is very similar to the location recommendation, many existing works [2],[3],[4] intuitively applied a location recommendation approach as their location prediction model. However, there are a few difficulties in adopting the location recommendation for the location prediction. First, the location recommendation process is a non-real-time estimation which means that the recent movements of the user are not taken into account in making the recommendations. Second, conventional location recommendation methods only consider the interest of the user such that they suggest a new location that a user may be interested in. However, the problem of next location prediction focuses on inferring the next location that a user will visit which not only considers the user's interest, but also the intention of the user. People do not solely visit locations because they are interested, they also go to places because they have to. Consequently, it is not straightforward to apply these recommendation techniques in location prediction.

On the other hand, considering the fact that human movement exhibits sequential patterns, various sequential pattern mining techniques [5],[6],[7],[8] have been developed for location predictions. These approaches address the location prediction as a historical movement matching problem. They usually consider the user's movement trajectory as a sequence of locations, and then, extract the frequent movement patterns from the set of trajectories. These frequent patterns then will be used as the prediction rules to be matched with the previous movement of the user. The difference of these approaches is

mainly about the type of the movement pattern they discover. However, they did not take the personalization into account as their approaches just return the same sequence patterns for all the users.

To extract the significant movement patterns, the existing methods mine the frequent sequences of locations from individual user trajectories, such that they assume all the movements as the solo movements. Accordingly, the extracted movement patterns only reflect the individual intention/interest. However, it has been shown that people do not visit locations just for their personal intention/interest. They also go to places which are motivated by the group's intention/interest they travel with [9], [10]. The preferences of the travel group, which may comprise very diverse people, have significant impacts on the places that users visit. Taking a family comprising two adults and a child who walking in a shopping mall as an example, considering only the individuals, i.e. parents, may lead to the different location prediction than when the group type, family, is taking into account.

In this paper, we propose a personalized framework to predict the next location of the users. The core idea underlying our proposal is the discovery of significant movement patterns by considering not only the individual movements, but also the group movements. We first, extract the groups of people who travel together, group travelers, from the spatial-temporal trajectories. Second, the profile information of the users will be used in order to identify the type of the extracted group. Third, the significant movement patterns will be discovered taking into account the group specific movements and individual movements. Finally, the problem of location prediction will be formulated as an estimation of the probability of a given user visiting a given location based on his/her current movement and his/her group type.

In order to extract the group travellers from spatial-temporal trajectories, we propose a novel group pattern, Loose Travelling Companion Pattern (LTCP), with taking into account the properties of human movement behaviour. Extracting the significant movement rules to support the prediction model is also a critical and challenging issue. We define two categories of significant movement rules: General sequential rules (SR) which refers to the movement rules considering the individual movement pattern, and Group-based sequential rules (GSR) which refers to movement patterns associated with the group types. Discovered movement rules then are utilized to construct the prediction model with further incorporating the distribution of places and group types.

The main contributions of this paper are summarized below:

- To the best of our knowledge, this is the first work that investigates the location prediction problem with consideration of group movement;
- We propose a novel group discovery approach to identify the groups of people who move together considering the human movement behaviour;
- We apply a classification model to predict the type of the discovered groups, utilizing the profile information of the users;
- We propose a novel location prediction model that takes into account the general movement rules and the group-based movement rules to predict next location of the user;
- We present comprehensive experimental results over various datasets. The results demonstrate that our proposed framework significantly outperforms the widely used sequential prediction technique.

The remainder of this paper is organized as follows. First, we briefly review the related work in Section II and present an overview of our proposed prediction framework in Section III. Next, our proposed group pattern discovery approach and group type prediction technique are described in detail through Section IV and V, respectively. The movement rules are discovered in Section VI, and the prediction model is constructed in Section VII. A real case study is introduced in Section VIII. The performance of our proposal through an empirical evaluation study is discussed in Section IX. Finally, conclusion and directions for the future works are given in Section X.

II. Related Work

The problem of predicting the future location has been variously formulated in the literature. The first strategy is Vector Based Prediction model which estimates the object's future location through applying the motion functions. These approaches can be divided into two types: 1) linear models assume that object's movement follows a linear pattern [11], [12], [13], [14], [15], and 2) non-linear models, on the other hand, take into consideration both linearity and non-linear patterns in modelling the object's movement [16], [17]. As the non-linear methods apply more sophisticated functions, they result in the higher prediction precision than the linear models. However, the motion functions are only able to predict the near future location. They also cannot differentiate between the random movement and regular movement of the object. These approaches are highly sensitive to the change in an object's movement, they cannot capture the sudden changes of the object as the function is only affected by the previous locations.

The movements of people contain a high level of regularity [18]. According to this fact, researchers have discovered the usefulness of extracting these regularities and applying them in order to predict the next movement of people. Accordingly, two prediction approaches have been raised: 1) discrete-time Markov model-based methods [19], [20], [21], [22], and 2) trajectory pattern based approaches [23], [24], [25], [26]. Markov model-based approaches extract a statistical method to estimate the next locations of the object among the spatial cells. They take which cell the object belongs currently into account, and then, calculate the next cell that the object is likely to be there in future. However, these approaches do not take the full movement history of the user into account. The new location depends on not only the last visited location but also on the previously visited locations.

The approaches mentioned above do not consider the influence of sequential visiting of locations on the user's

movement behaviour, although, in reality, human movement exhibits sequential patterns [18], [27]. Trajectory pattern based prediction approaches address the location prediction as a problem of matching the historical movement of the user on the extracted frequent movement patterns. Accordingly, prediction techniques developed for this problem domain can be broken down into two steps: 1) extracting the frequent movement patterns, and 2) prediction model building.

Frequent movement pattern Extraction : Various techniques have been developed to extract the frequent patterns from movements of users [23], [28]. These methods can be classified according to what movement information they use to extract the patterns, for example considering the location only, time, or semantic information. The first group of approaches consider a trajectory as a sequence of locations, and then use the existing frequent pattern mining approaches to extract the movement patterns [5], [6], [7], [29]. In [5] and [6] several methods have been proposed to generate the association rules for an individual user using a modified versions of the Apriori algorithm [30]. Such rules can identify the frequent locations which are visited together in the movement of an individual user. In order to choose the appropriate rule for the prediction, they take two criteria of support and confidence into consideration. Morzy et al. [7] subsequently applied a modified version of the PrefixSpan algorithm [31] to discover the sequential frequent movements of users. Such pattern can identify the locations which have been visited together, along with the consequence of location, i.e., the place users mostly visit after visiting somewhere else. Jeung et al. [32] proposed an innovative approach that combine a vector-based model with the trajectory based model to forecast the future locations of a user. They apply Recursive Motion Function [16] to predict the near future locations, and to discover mobile sequential patterns a modified version of the Apriori algorithm [30] is used.

Spatial-temporal data may reveal various discoveries such as location pattern recoginition and networks conditions [33]. By considering the temporal information of the movement besides the spatial information, spatial-temporal sequential patterns can be extracted. Giannotti et al. [34] proposed T-pattern, a kind of spatial-temporal sequential pattern, which takes into consideration both spatial and temporal information of user's movement. In order to identify the temporal patterns, temporal information is mapped in the R^n space and then the dense hypercubes are discovered from the R^n space. In [23], the authors extended their previous work by proposing three mining algorithms to extract the frequent movement patterns from the spatial-temporal trajectory data. The proposed algorithms are different in terms of how they define the stay locations. In [8] and [35] the authors propose a pattern which takes into account the semantic information in addition to the spatial-temporal information to extract the frequent movement patterns of users. The idea of taking service requests into account, mobile access pattern, was initially raised by Tseng et al. in [36]. This pattern considers the user's movement along with the associated service requests in each location. In [37] and [38] an efficient approach for mining the sequential mobile access patterns from users' movement was proposed, SMAP-Mine, which is based on the FP-Tree [39]. T-MAP [40] was proposed by Lee to efficiently identify the mobile user's access patterns taking the time intervals in addition to the location and service information into account. Yun et al. in [41] proposed the Mobile Sequential Pattern (MSP) which is taking the moving paths in mining the frequent patterns. The existing studies, however, can not differentiate movement behaviours among users. The reason is, they focus on discovering the movement patterns from the whole movement data. To alleviate this problem, Lu et al. in [42] proposed a novel pattern, Cluster-based Temporal Mobile Sequential Pattern, which takes different clusters of users into account. However, their clustering method is based on the movement trajectory of users, not the users' personal information. Consequently, the prediction model can not deal with the problem of personalized prediction of the next location motivated by group intentions.

Prediction model building : Existing studies on user location prediction can also be classified according to the reference data they use for building the prediction model: 1) those who their model is only based on the movement history of the user itself, 2) those who build the model using the movement data of all the users in a database, and 3) hybrid approaches who take advantage of both kinds of data. First category of studies model the regular movement of a user in order to predict his/her next location [5], [32]. However, as the historical movements of all users are not included in the prediction model, it results in low coverage of prediction. Even if a user has visited many locations, there are places that the user has never been there. As a result, the locations not previously seen by the user, can not be predicted with this prediction model. The second category of studies, apply only the movement data from all users to predict the next location, like probability distribution based models [43], or location recommendation models [2], [3], [4]. However, as these kinds of recommenders do not take the current movement of the user into account, they result in the low precision prediction. The third category of studies predicts next location of users using a hybrid method, which not only consider a user's current movement data but also utilize the movement data from other users [44]. Even though, this branch of studies alleviates the low coverage and low precision of the tow above categories, it still suffers from the following problem: as the prediction model focuses only on solo movements of users, it can not deal with the next location prediction motivated by the group-triggered intentions. Consequently, the prediction model is not able the predict the next location motivated by the group intentions.

III. System Overview

Fig. 1 shows the architecture of our system, which is divided into two main parts, "Offline" and "Online". In the offline part, the prediction rules are minded from the existing movement trajectories, and through the online part, the location prediction for the incoming trajectory will be done. The groups of people

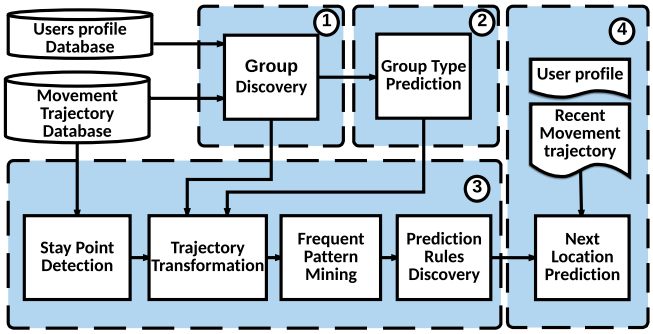


Fig. 1. System Architecture

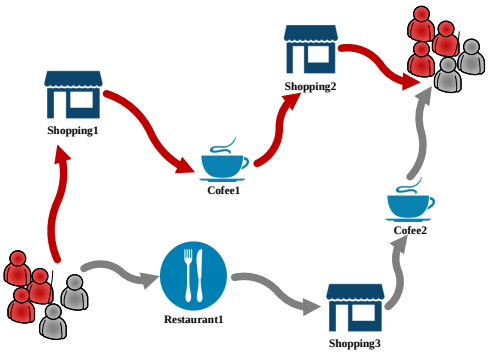


Fig. 2. A group of passengers at airport

are firstly detected from the movement trajectories (Section IV). The individual information of group members then is used for conducting the travel group type prediction (Section V). The stay point locations are extracted from the spatial trajectories and the location sequences which are associated with the travel group types will be generated. Through applying the sequential pattern mining the significant movement patterns will be identified from the generated location sequences, and then, prediction rule will be extracted (Section VI). Finally, we propose a location prediction model, which is entailed by the general sequential prediction rules and the group-specific sequential prediction rules (Section VII).

A. Input Data Characteristics

The input data we are working with in this paper, consists of two pars: movement trajectories, and the profile information.

$Movement\ Trajectory$: $O_{DB} = \{O_1, O_2, ..., O_n\}$ is a set of people in our database. The spatial-temporal trajectory Tr of each person O is represented as a series of points denoted as $Tr = \langle p_1, p_2, ..., p_n \rangle$, where p_i includes the location and timestamp attributes.

$Profile\ Information$: Each person is also attached with some individual information $Id = \langle age, gender \rangle$. Therefore, for each person we have the following pair of information $O_i = \langle Tr_i, Id_i \rangle$. In this paper, we put this assumption that we have access to the individual information of the people. These information can be identified from various ways, i.e., such as extracting the profile information from the users' social network or or as the input data from an application. Addressing these details is out of scope of this paper. In section VIII, we provide a realistic example which provide us with this kind of information.

IV. Group Discovery from People Movement Trajectories

Several studies have been proposed in the literature to discover the groups of moving objects. The main focus of the previous approaches, however, is on the movement trajectories of vehicles or animals with the aim of finding the general trends [45], [46]. Different from them, we concentrate on human movement trajectories, particularly in the indoor environment. The movement of people is rather different from the movement of animals or vehicles. People might belong to the same group, while they have different movements and follow the different path. Groups evolve and form various sub-groups during their lifetime. Fig.2 represents an example of group movement of people browsing the airport before their departure (according to the observation of movement data of passengers at Guangzhou Baiyun International Airport). As it is shown in the figure, while these passengers belong to the same main group, they also contribute in different sub-groups.

Considering the mentioned fact, we propose a novel group discovery approach, Loose Traveling Companion Pattern (LTCP) framework, to identify the group travellers including the main group and the sub-groups they form during their movement [47]. In the following, first, the essential concepts will be explained, afterward, we present the group discovery problem in a formal way. The list of main notations is outlined in Table I.

A. Essencial Concepts

We consider $O_{DB} = \{O_1, O_2, ..., O_n\}$ as a set of moving objects (corresponding to the people in our problem). The spatial-temporal trajectory Tr of an object O is represented as a series of points denoted as $Tr = \langle p_1, p_2, ..., p_n \rangle$, where p_i includes the location and timestamp attributes. We assume $t \in \{1, ..., T\}$ as the time interval, which T may be equal to a day or a time slot can be determined by 1 minute, depending on the requirement of applications. The set of objects with their trajectories at time slot t is called a **slot-dataset** at t.

Then we extract the clusters of objects at each time slot t, **slot-cluster**, according to the corresponding $slot - dataset$. The output of this step, is a database of slot-clusters $C_{DB} = \langle C_1, C_2, ..., C_n \rangle$. Each slot-cluster C_i contains the extracted clusters at time-slot i, $C_i = \langle c_{i,1}, c_{i,2}, ..., c_{i,j} \rangle$, where j is the number of clusters at that time-slot. The number of time-slots that a cluster c has been observed is denoted by $c.f$. We call the subset set of clusters at time-slot i, **subset-collection** $S_i = \langle cs_{i,1}, cs_{i,2}, ..., cs_{i,k} \rangle$, where k is the number of subsets (except empty subset). Each subset cs is called a **cluster-set**. For example, in Fig. 3, at time-slot 2, we have the slot-cluster $C_2 = \langle c_2, c_3 \rangle$ and subset-collection $S_2 = \langle \langle c_2 \rangle, \langle c_3 \rangle, \langle c_2, c_3 \rangle \rangle$,

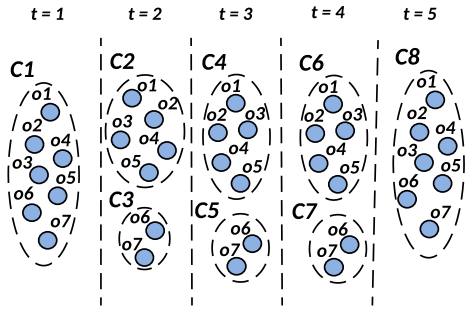


Fig. 3. Example of movement pattern

and the cluster-set $cs_{2,3} = \langle c_2, c_3 \rangle$. The timestamp of a cluster-set is denoted as $cs.t$.

B. Problem Formulation

An LTCP group is a sequence of cluster-sets at continuous time-slots. We define four threshold parameters in order to identify the LTCP groups: 1) size threshold m_G which restricts the size of the groups we are targeting, 2) duration threshold d_G which determines the minimum lifetime of a group, 3) frequency threshold f_C that determines the minimum time-slots that a group should be gathered.

$Definition\ 1$: According to the above mentioned parameters, we identify an object set O_G along with a cluster-set sequence $TS = \langle cs_{1,a}, cs_{2,a2}, ..., cs_{n,an} \rangle$ $(n = t_2 - t_1 + 1)$ at interval $[t_1, t_2]$ as an LTCP group, if it satisfies the following conditions:

1) $cs_1.t = t_1$, $cs_n.t = t_2$, $|t_2 - t_1| \geq d_G$;
2) $|O_G| \geq m_G$;
3) For any $cs_{i,ai}$ of TS, union of its clusters should be equal to O_G : $\forall\ c_j \in cs_{i,ai}$, $\bigcup c_j = O_G$;
4) $\exists\ c_j \in TS$, $c_j = O_G$, and $c_j.f \geq f_C$;

then the pair of object set O_G and cluster-set sequence TS is defined as **Loose Travelling Companion Pattern (LTCP)** in $[t_1, t_2]$, $P = \langle O_G, TS \rangle$. Based on the union operation among clusters in a cluster-set, a sequence of cluster-sets is derived which meets condition 3. Condition 4 adds a constraint on the number of time-slots that all members of an LTCP group should stay close together. Fig. 3 shows an example of LTCP group with members $\{O_1, O_2, O_3, O_4, O_5, O_6, O_7\}$ in interval [1, 5] while it satisfies the condition of $f_C = 2$ and $d_G = 5$.

However, the rigid continuous time constraint in LTCP may lead to no discovery of a group or fragmented discovery of a group. In cases with lots of objects in a limited space, i.e., airport, it might be that the objects in a group stay in the same cluster with other groups for a few time-slots. Strict continuous time constraint in LTCP will prevent the discovery of such group patterns. Therefore, we propose **Weakly Continuous Loose Travelling Companion Pattern (WCLTCP)**, which is the extension of LTCP. The difference between WCLTCP and LTCP is the possibility of a time-gap between the cluster-sets. Considering cluster-set sequence $TS = \langle cs_{1,a1}, cs_{2,a2}, ..., cs_{n,an} \rangle$ and time-gap threshold l_C, the following condition should be satisfied in a WCLTCP group: $cs_{i+1}.t - cs_i.t \leq l_C$ $(\forall\ i,\ 1 \leq i < n)$.

Example: Table II shows the running process of the LTCP discovery algorithm. It is clear that the membership is unchanged during the lifetime of the group (O_G = $\{O_1 - O_9\}$); however members are contributing in different sub-groups $\{O_1 - O_5\}$ and $\{O_6 - O_7\}$. In the following sections, we explain how the discovered groups (main-group and sub-groups) are used to mine the significant movement patterns.

As the focus of this paper is on the prediction part, we omitted the explanation of the discovery algorithms. The detailed about the discovery algorithms of the proposed patterns, along with the extensive evaluations, has been provided in [47].

TABLE I
TABLE OF NOTATIONS

Notation	Definition
O	Moving object
Tr	Trajectory
t	Time slot index
O_{DB}	Set of objects
c	Cluster
$c.f$	Frequency of a cluster
C_i	Set of clusters at time slot i, slot-cluster
C_{DB}	Set of slot-clusters at all time slots
cs	Subset of C_i, cluster-set
S_i	Set of all subsets of C_i, subset-collection
$cs.t$	Timestamp of a cluster-set
m_G, d_G	Size and duration threshold
f_C	Frequency threshold of gathering of all members
l_C	Gap threshold between cluster-sets

V. PREDICTING TRAVEL GROUP TYPES

It has been pointed out that the preferences of the group people travel with have a significant impact on the locations they visit. Taking a family group and a couple group as examples, predicting a location preferred by the family to both groups may not satisfy the couple preferences.

Through the previous step, we discovered the traveller groups from the movement trajectory data. In this step, we identify the type of the discovered groups by utilizing the profile information, Id, of the users. In this paper, we take four

TABLE II
ILLUSTRATION OF LTCP DISCOVERY

time	cluster-set	candidates
1	C_1	$P_1 = \langle\ O_G, \langle\langle C_1 \rangle\rangle\ \rangle$
2	C_2	$P_3 = \langle\ O_G, \langle\langle C_1 \rangle, \langle C_2, C_3 \rangle\rangle\ \rangle$
3	C_4,C_5	$P_3 = \langle\ O_G, \langle\langle C_1 \rangle, \langle C_2, C_3 \rangle, \langle C_4, C_5 \rangle\rangle\ \rangle$
4	C_6,C_7	$P_4 = \langle\ O_G, \langle\langle C_1 \rangle, \langle C_2, C_3 \rangle, \langle C_4, C_5 \rangle, \langle C_6, C_7 \rangle\rangle\ \rangle$
5	C_8	$P_5 = \langle\ O_G, \langle\langle C_1 \rangle, \langle C_2, C_3 \rangle, \langle C_4, C_5 \rangle, \langle C_6, C_7 \rangle, \langle C_8 \rangle\rangle\ \rangle$

group types into account, (1) family, (2) friends, (3) couple and (4) solo traveler, as they have been shown to have a significant impact on choosing the location to visit [9]. In the following, the required features for representing the group travellers will be introduced and the prediction model will be trained by means of these features.

Considering the profile information of the group members (age and gender), we identify the influential features on type of the group travellers. For example, a family group usually includes parents (gender difference) and one or more kids (age gap), or a couple group includes a male and female (gender difference), who are usually close in age (age gap). In detail, three kinds of features are used in the proposed group type prediction method.

1) Gender difference: By averaging the L2 distance of gender between any two persons, the gender difference will be calculated.
2) Age gap: We consider two types of age-based measures, (1) age-gap, and (2) age range. The first one is calculated by the standard deviation and the second one is the average of the members' age.
3) Group size: The number of members in a group is another influential parameter, for example, group of friends or family groups are larger than the couple groups.

Given a set of traveller groups and the corresponding travel group types, we will train the prediction model. In order to build the prediction model, we apply the Adaboost where the weight and the threshold of each feature is determined. The similar approach has been used in [9].

In this paper, we consider groups with one member as a solo traveler, therefore, we would train our prediction model on data from three other group types (family, couple, and friends). We will demonstrate the effectiveness of group type prediction and its impact on location prediction in Section IX.

VI. Group-Specific Pattern Mining

Through the previous sections, traveller groups and their types are discovered. In this section, we use this information to discover the significant movement patterns, the patterns which are frequently visited by the users. We propose a new type of pattern, called GFSP (Group-Based Frequent Sequential Pattern), to represent the frequent movement behaviours by considering the group travellers. In contrast to the conventional sequential pattern, which considers the individual movements, we take into account the type of the group to illustrate the movement of users. As shown in Fig. 1, we first detect stay locations from spatial-temporal trajectories; then we transform each trajectory to a sequence (or sequences) of stay locations, considering the discovered groups, called GLS (Group-based location sequence). Afterward, we apply a frequent sequential pattern mining to discover GFSPs from GLSs, and then extract the prediction rules. The list of main Notations are outlined in Table III.

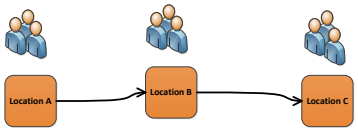


Fig. 4. Simple group movement

A. Stay Location Discovery

Before identifying the significant movement patterns, the raw spatial-temporal trajectories need to be transformed to the sequences of stay locations visited by the users.

Stay location is defined as the geographic region user stay for over a time threshold. In this phase, we follow a grid based approach which use a regular grid (or some pre-defined spatial decomposition) to divide the space into cells, locations. The time a user stays in a cell, stay time, is obtained according the difference between the time a user enters and leaves the cell. Cells with stay time shorter than the user-specified time threshold, st_U, will be filtered out. We call the remaining cells (i.e., their stay time is equal to or greater than the threshold) stay cells, stay locations. It should be noted, various approaches can be applied for identifying the stay location depending on the input data.

B. Trajectory Transformation

Intuitively, after stay locations are detected, trajectories can be transformed to a sequence of stay locations. Applying the conventional transformation approach, which considers individual trajectories, we can transform each trajectory into a sequence of stay location. However, in group-based location sequence transformation, we consider the group information, rather than the individual information, to transform the spatial-temporal trajectories into the location sequences.

Fig.4 represents a simple movement of a group (i.e., family) which members follow the same path during the whole life time of the group. Subsequently, conventional approach generates three location sequences (LS) considering the individual movements:

$$LS\ 1 : A \rightarrow B \rightarrow C$$
$$LS\ 2 : A \rightarrow B \rightarrow C$$

TABLE III
Table of Notations

Notation	Definition
GSP	Group-Based Sequential Pattern
LS	location sequence
GLS	Group-based location sequence
FSP	Frequent sequential pattern
$GFSP$	Group-based frequent sequential pattern
SR	Sequential Rule
GSR	Group-based sequential Rule

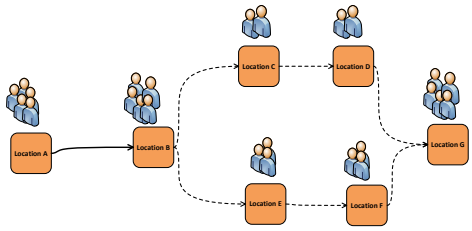


Fig. 5. Complex group movement

$LS\ 3 : A \rightarrow B \rightarrow C$

Unlike the conventional approach, group-based location sequence transformation considers groups rather than the individuals. Consequently, group-based transformation generates one group-based location sequence (GLS) considering the group movement.

$GLS\ 1 : family(A \rightarrow B \rightarrow C)$

Fig. 5 shows another example of movement of a group travellers, friends, which is composed of two sub-groups, a family and a couple. They (friends) all visit Location A and Location B, then first sub-group (couple) visits Location C and Location D and the second sub-group (family) visits Location E and Location F. They eventually join in Location G. Applying the conventional transformation, movement trajectories are turned into the following location sequences:

$LS\ 1 : A \rightarrow B \rightarrow C \rightarrow D \rightarrow G$
$LS\ 2 : A \rightarrow B \rightarrow C \rightarrow D \rightarrow G$
$LS\ 3 : A \rightarrow B \rightarrow E \rightarrow F \rightarrow G$
$LS\ 4 : A \rightarrow B \rightarrow E \rightarrow F \rightarrow G$
$LS\ 5 : A \rightarrow B \rightarrow E \rightarrow F \rightarrow G$

Unlike the conventional approach, which each trajectory is transformed exactly to one location sequence, group-based location sequence transformation doesn't follow the same straight-forward approach. We have three group types in the above example, friends (main-group), couple (sub-group), and family (sub-group). Subsequently, trajectories are transformed to the group-based location sequences considering their group types:

$GLS\ 1 : friends(A \rightarrow B \rightarrow C \rightarrow D \rightarrow G)$
$GLS\ 2 : friends(A \rightarrow B \rightarrow E \rightarrow F \rightarrow G)$
$GLS\ 3 : couple(A \rightarrow B \rightarrow C \rightarrow D \rightarrow G)$
$GLS\ 4 : family(A \rightarrow B \rightarrow E \rightarrow F \rightarrow G)$

These location sequences then are treated as the reference data for training and prediction.

Eventually, two kinds of databases containing the location sequences are extracted from the movement trajectories, general database and group-specific database. General location sequence database, $D_{General}$, is obtained through applying the conventional transformation method, with focusing on the individuals. This database contains the location sequences for the corresponding users. On the other hand, Group-specific location sequence database, $D_{Group-specific}$, is obtained through applying the group-based transformation with considering the group information. This database contains location sequences belonging to the corresponding group types.

TABLE IV
SAMPLE DATABASE

User ID	Location Sequence	Group Type
1	$A \rightarrow B \rightarrow C \rightarrow D$	$Family$
2	$A \rightarrow B \rightarrow C \rightarrow D$	
3	$A \rightarrow B \rightarrow C \rightarrow D$	
4	$A \rightarrow B \rightarrow C \rightarrow D$	$Family$
5	$A \rightarrow B \rightarrow C \rightarrow D$	
6	$A \rightarrow B \rightarrow C \rightarrow D$	
7	$E \rightarrow F \rightarrow G$	$Couple$
8	$E \rightarrow F \rightarrow G$	
9	$H \rightarrow I \rightarrow J$	$Couple$
10	$H \rightarrow I \rightarrow J$	
11	$E \rightarrow F \rightarrow G$	$Solo$
12	$H \rightarrow I \rightarrow J$	$Solo$

C. Identifying the Significant Movement Patterns

During the two previous phases of our four phases prediction framework, we extracted the location sequences (LS and GLS) from the movement trajectories. In the third phase, frequent pattens are mined from the location sequences, and then, prediction rules are extracted from these patterns.

To extract the frequent sequential patterns, we apply the widely used Apriori algorithm in [30]. We define support of a pattern, as the number of times it has been visited by the users, $supp$. A pattern is frequent, if its support is higher than the specified support threshold, δ. Accordingly, first, we mine the frequent sequential patterns (FSP) from $D_{General}$. Group-based frequent sequential patterns (GFSP), then, are extracted from the group-based location sequences in $D_{Group-specific}$. Table IV shows a sample database of location sequences. Table V and Table VII are the corresponding General database and Group-specific database, respectively. Considering $\delta = 2$, FSPs and GFSPs are mined in Table VI and VIII, respectively.

D. Sequential Rules

After discovering the frequent patterns, we generate the sequential rules (SRs) and group-specific sequential rules (GSRs) from the FSPs and GFSPs, respectively.

For a given pattern $FSP_L : \langle L_1, L_2, ..., L_n \rangle$, sequential rule SR_L and the confidence $Conf(SR_L)$ are written as:

$$SR_L = \langle L_1, L_2, ..., L_{n-1} \rangle \rightarrow \langle L_n \rangle \tag{1}$$

$$Conf(SR_L) = \frac{Sup(\langle L_1, L_2, ..., L_n \rangle)}{Sup(\langle L_1, L_2, ..., L_{n-1} \rangle)} \tag{2}$$

TABLE V
GENERAL DATABASE ($D_{General}$)

User ID	Location Sequence
1	$A \rightarrow B \rightarrow C \rightarrow D$
2	$A \rightarrow B \rightarrow C \rightarrow D$
3	$A \rightarrow B \rightarrow C \rightarrow D$
4	$A \rightarrow B \rightarrow C \rightarrow D$
5	$A \rightarrow B \rightarrow C \rightarrow D$
6	$A \rightarrow B \rightarrow C \rightarrow D$
7	$E \rightarrow F \rightarrow G$
8	$E \rightarrow F \rightarrow G$
9	$H \rightarrow I \rightarrow J$
10	$H \rightarrow I \rightarrow J$
11	$E \rightarrow F \rightarrow G$
12	$H \rightarrow I \rightarrow J$

TABLE VI
FREQUENT SEQUENTIAL PATTERNS (FSP)

Pattern	Support
$A \rightarrow B \rightarrow C \rightarrow D$	6
$E \rightarrow F \rightarrow G$	3
$H \rightarrow I \rightarrow J$	3

TABLE VII
GROUP-SPECIFIC DATABASE ($D_{Group-specific}$)

Location Sequence	Group Type
$A \rightarrow B \rightarrow C \rightarrow D$	$Family$
$A \rightarrow B \rightarrow C \rightarrow D$	
$E \rightarrow F \rightarrow G$	$Couple$
$H \rightarrow I \rightarrow J$	
$E \rightarrow F \rightarrow G$	$Solo$
$H \rightarrow I \rightarrow J$	

TABLE VIII
GROUP-SPECIFIC FREQUENT SEQUENTIAL PATTERNS (GFSP)

Pattern	Group Type	Support
$A \rightarrow B \rightarrow C \rightarrow D$	$Family$	2

In the definition of confidence of SR_L, we term the antecedent $\langle L_1, L_2, ..., L_{n-1} \rangle$ as left hand side (LHS) and the consequent $\langle L_n \rangle$ as right hand side (RHS).

In order to reveal the strength of each rule, we rank the rules by considering both support and confidence of the rules:

$$Strength(SR_L) = Sup(\langle L_1, L_2, ..., L_n \rangle) \times Conf(SR_L) \tag{3}$$

The same procedure will be applied for extracting the GSRs.

VII. GROUP-SPECIFIC LOCATION PREDICTION MODEL

According to Fig. 1, the input of the prediction framework include the target user's profile composed of the group he/she belongs, historical movements (sequence of the visited locations), and the output is the most probable location he/she is going to visit. First, the group that the user belongs to is discovered through applying the group discovery approach, as described in Section IV. Then, the type of the group (family, friends, couple, or solo traveler) will be obtained by applying the proposed group type prediction model. Afterward, the known trajectory of the user will be compared with the movement prediction rules generated from the frequent trajectories (SRs and GSRs).

A. Prediction Model

In this section, we describe how the discovered rules (SRs and GSRs) are applied to predict the next location of the user according to the user's historical movement. let $S_U = (s_1, s_2, ..., s_m)$ be a trajectory of a user, for which we are seeking the most probable next location. For a given trajectory S, set of all matched rules is denoted by $R_S = (r_1, r_2, ..., r_n)$. For a given user's trajectory S_U, we call rule $r_i = \langle LHS_i \rangle \rightarrow \langle RHS_i \rangle$ a matched rule, such that LHS_i covers a part of the trajectory S_U, $LHS_i \subseteq S_U$. The strategy outputs, as the result, the rule's right hand RHS_i, weighted by the relative coverage of the trajectory S_U. For a given movement rule r_i the score of the prediction is defined as:

$$Score(r_i, S_U) = Strength(r_i) \times \frac{length(LHS_i)}{length(S_U)} \tag{4}$$

the strength of the rule is calculated according to $Equation$ 3.

To predict the optimal next location, this problem can be formulated as follows:

$$L^{Next} = argmax_{L_J} UScore(L_J | S_U, G_U, SR, GSR) \tag{5}$$

where G_U is the learned group type for the corresponding user, and SR and GSR are the obtained sequential rules. Intuitively, we will predict the proper location that most users visit and match the user's group type and historical movements S_U.

Until now, we have two kinds of rules: the general sequential rules $SR_i = \langle L_1, L_2, ..., L_{n-1} \rangle \rightarrow \langle L_n \rangle$ which show the popularity of the different location sequences. The group-specific sequential rules $GSR_i = (\langle L_1, L_2, ..., L_{n-1} \rangle \rightarrow \langle L_n \rangle, G_T)$ which show the popularity of the different location sequences for each group type G_T. Here we describe a way of constructing the scoring model that takes two rule sets into account:

$$UScore(L_J | S_U, G_U, SR, GSR) = UScore(L_J, S_U, SR) \cdot UScore(L_J, S_U, G_U, GSR) \tag{6}$$

where $UScore(L_J, S_U, SR)$ calculates the unified score of the location L_J according to the general sequential rules, and

$UScore(L_J, S_U, G_U, GSR)$ calculates the unified score of the location L_J according to the group-specific sequential rules as follows:

$$UScore(L_J, S_U, SR) = \sum_{r_i \in SR, RHS(r_i)=L_J} Score(r_i, S_U) \tag{7}$$

$$UScore(L_J, S_U, G_U, GSR) = \sum_{r_i \in GSR, RHS(r_i)=L_J} Score(r_i, S_U) \tag{8}$$

where $Score(r_i, S_U)$ is calculated by $Equation$ 4. It should be noted, here we applied the general rules and group-specific rules with the same importance. However, different weights can be assigned according to the goal of the application.

B. Background Smoothing

There is a possibility that we cannot find a matched rule in any of the rule sets (SRs and GSRs) for the incoming trajectory. For example, before we would be able to discover the group of a target user, we cannot utilize the GSRs to predict the next location, and we cannot predict the next location, according to $Equation$ 6. In order to solve this problem, we need to add a smoothing factor in the scoring model. Intuitively, we take location popularity and group-specific popularity into consideration:

$$Popularity(L_J) = \frac{Frequency(L_J)}{TotalNumberofVisits} \tag{9}$$

$$Popularity(G_U) = \frac{Frequency(G_U)}{TotalNumberofGroups} \tag{10}$$

where $Frequency(L_J)$ refers to the number of visiting the location L_J and $TotalNumberofVisits$ refers to the total number of visiting by all the users. Similarly, $Frequency(G_U)$ refers to the number of groups with type of G_U to the total number of discovered groups $TotalNumberofGroups$.

Combining the scoring model with the popularity information, $Equation$ 9 and $Equation$ 10 into $Equation$ 6, we get the following model as the final prediction model:

$$L^{Next} = argmax_{L_J} \{Popularity(L_J) + UScore(L_J, S_U, SR)\} \cdot \{Popularity(G_U) + UScore(L_J, S_U, G_U, GSR)\} \tag{11}$$

Because the term $Popularity(L_J) + UScore(L_J, S_U, SR)$ considers the relationship between locations by utilizing the general sequential rules and location popularity, the prediction using this is regarded as $general\ sequential\ prediction\ model\ (SPM)$. On the other hand, the term $Popularity(G_U) + UScore(L_J, S_U, G_U, GSR)$ in the prediction is the $group-based\ sequential\ prediction\ model\ (GSPM)$. The combination of SPM and $GSPM$ is denoted as "$SPM + GSPM$".

Fig. 6. Charlie/ Smart Trolley

VIII. Charlie: Smart Trolley

Over the last decade, airport industry has evolved dramatically to a commercial enterprise with the focus on the importance of customers. In large airports, like Heathrow Airport in London, thousands of passengers are served every day. Providing the right services to the passengers is the ultimate goal for the airport authority. In this regard, access to the passenger information plays an important role.

Charlie is a smart trolley developed by working with industrial partner Wuxi Chigoo Interactive Technology Co. Ltd in China, Fig. 6. This tool is a combination of a trolley and Android based tablet. The goal of Charlie is helping passengers to navigate inside the airport and provide them with personalized services. By using the smart Charlie, passengers may carry their handbags, receive personalized flight information on the tablet, receive boarding reminding, enjoy the media and Internet services, and navigate themselves inside the airport.

To begin, the passenger needs to scan his/her boarding pass on the Charlie. This provides us the access to the profile information of the passenger, i.e. Flight Number, Destination, Age, and Gender. Charlie, in addition, reports its location which capables us to track the passenger. The location data is reported with the format of: $\langle MacAddress, Coordinate, Time \rangle$. Through the first field, we can distinguish the Charlies, and through the other two fields we have access to the movement trajectory of the passengers. In this paper, we utilize the movement information to extract the passenger groups and the individual information, Age and Gender, will then be used to identify the type of the discovered groups. More detailed information can be found in [47].

IX. Experiments

In this section, we conducted a series of experiments to evaluate the performance of the proposed prediction model, under various conditions. Experiments can be divided into three parts: 1) group discovery evaluation, 2) group type prediction evaluation, and 3) next location prediction evaluation.

All of the experiments were implemented in Python on a 3.30 GHz machine with 4 GB of memory running Ubuntu.

A. Datasets

As mentioned in Section VIII, we use our smart trolley, Carlie, to collect the profile information and also the location information of the passengers. In order to obtain a comprehensive dataset for evaluating our model, we build a framework which utilize both the real and synthetic datasets.

1) dataset 1 (D_1): To be able to evaluate the accuracy of the proposed group discovery approach, we conducted an experiment to obtain the ground truth. The experiment took place in Guangzhou Baiyun Airport with 100 participants who used Charlie to browse the airport and report their location. During the experiment, participants were divided into groups with different sizes between 1-8. Each group followed the predefined routes which were chosen such that the members split and merge and stop for several times. The groups spent between 1 to 2 hours browsing the airport. In order to obtain a more complete dataset comprising more groups with different sizes and lifetimes, the period that the passenger spends at airport, we generate a number of groups according to the real movement data received from the experiment. The resulted dataset is comprised of 5000 passengers which belong to the groups with different sizes of 1 to 8 and lifetime between 1 to 6 hours.

2) dataset 2 (D_2): In order to obtain the training data for predicting group types, we interviewed passengers at airport (Guangzhou Baiyun Airport during the July 2017) about the type of the group they travel with (family, couple, friends, or solo traveler) and the individual information (age and gender) of the travelers. Eventually, around 1000 group information was collected (more details are listed in Table IX). Group information contains the corresponding group type and the individual information of the members as follows: $\langle GroupType, \langle Age_1, Gender_1\rangle, ..., \langle Age_m, Gender_m\rangle\rangle$.

As it is mentioned before (Section V), we consider groups with one member as the solo travelers. Therefore, we only use the data from three other group types (family, couple, and friends) to train the group type prediction model.

3) dataset 3 (D_3): Currently, the only way for collecting the location data of passengers is through the Charlie. As not all members of a group use a separate trolley, we don't have the individual information and also the movement trajectory data for all the group members. For this reason, we propose a framework to build a comprehensive synthetic dataset with the help of the collected real datasets (D_1 and D_2) to evaluate the proposed location prediction model. This framework follows three major steps: 1) initializing the groups with different sizes and types, 2) generating the movement path for a group such that it reflects the influence of group type on the movement (our main assumption), and 3) generating the individual movement trajectories of passengers considering their group movement.

TABLE IX
TRAINING DATA FOR PREDICTING THE GROUP TYPE

	Family	**Couple**	**Friends**	**Solo-Traveler**
Collection	281	215	174	330

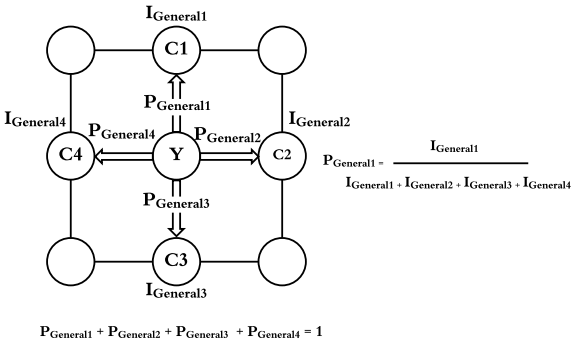


Fig. 7. Cell network for modelling the movement behaviour

In this experiment, we build a database containing N groups, $G = \{g_1, g_2, ..., g_N\}$, which are equally divided into four categories, family, couple, friends, and solo travellers. For each group, g_i, then we randomly choose an instance with the same group type from database D_2. For example, if type of g_i is a family, we choose a family group from D_2 and assign the corresponding group information to g_i, i.e. $g_i = \langle type : family, membersInfo : \langle\langle 56, Female\rangle, \langle 58, Male\rangle, \langle 16, Female\rangle\rangle\rangle$.

We model the environment as a $|W| \times |W|$ cell network. Cells are considered as the locations that passengers visit and have predefined length $|l|$ and width $|w|$. Each location has different level of interestingness for different group types. For example, a location which is not family favourite, could be interesting for couples. For this reason, we assign two values of interestingness to each cell, group specific interestingness $I_{Group-Specific}$ and general interestingness $I_{General}$. The former refers to the interestingness of a cell for a specified group type, i.e. I_{family}. The later points to the interestingness of a cell in general. The value of interestingness parameters are determined from a uniform distribution within a given range U_I. Eventually, each cell is assigned with a few interestingness values, I_{Family}, I_{Couple}, $I_{Friends}$, I_{solo}, and $I_{General}$.

For generating the movement path (movement between cells) for each group, we apply a modified version of simulation model in [41], [42]. For each cell, the advancing probability of each neighbour is the probability for a group to move to the neighbouring cells. In the model, the advancing probability p_a is obtained by the ratio of the interestingness value of each neighbour to those values of other neighbours. Corresponding to the different types of interestingness values, we define different types of advancing probabilities, group-specific advancing probability $P_{Group-Specific}$ which includes P_{Family}, P_{Couple}, $P_{Friends}$, and P_{Solo}, and general advancing probability $P_{General}$. The backward moving represents that a user will move from the current cell back to the cell

from which the user came. The backward probability p_b is denoted by $P_b = P_a.W_b$, where W_b is a backward weight. Similar to the advancing probability there are different types of backward probability. Each group may choose the next cell to move according to the group-specific advancing probability with probability P_{Event} or following the general advancing probability. Length of the movement path is determined from a Poisson distribution with a mean equal to U_T, respectively.

$Example$: For the 3×3 mesh network example shown in Fig. 7, there are four neighbours $\langle C_1, C_2, C_3, C_4 \rangle$ for cell Y. As an example, we assigned each cell with the general interestingness value, $I_{General1}$. When a user wants to move out from cell Y, the general advancing probabilities to the neighbours are shown in Fig 7. If a user has already came from one of the neighbours to the current cell, the advancing probability of the corresponding node should be calculated considering the backward weight.

So far, the groups and the corresponding movement path between cells are extracted. At the final step, we generate the movement trajectories of the individual members of the group inside and between the cells through the following procedure. We choose one of the members as the head who moves following the extracted group movement path. Other members, then, move approximately towards the head, similar to the group movement behaviour in D_1, splitting and merging during the movement. We define N_{tp} time points in a day and choose the start time point for each user trajectory from that. The movement data for each user, then, is generated every 10 seconds in the form of $\langle x, y, t \rangle$ which corresponds to the spatial coordinates and temporal information of the movement. The spending time in each cell (stay time) is determined from a Poisson distribution with a mean equaling to N_T.

The final dataset contains the spatial-temporal trajectories of users along with the users' individual information (age and gender) and the information of the group they travel with (group members and the group type). The default values for the above framework is listed in Table X.

B. Parameter Settings and Measurements

Default values of the parameters for discovering the groups are listed in Table XI. Detailed information about this step can be found in [47]. To extract the significant movement patterns, we set the stay time threshold as 10 minutes, st_U, and support threshold, δ, as 0.1%, Table XII. It should be noted that all of the parameters can vary according to the different applications. In this paper, we set the parameters according to our application which is predicting the next location for the passengers at airports.

The followings are the main predictability measurements for the location prediction evaluation:

$$Precision = \frac{P^+}{P^+ + P^-} \tag{12}$$

$$Recall = \frac{P^+}{|R|} \tag{13}$$

$$Applicability = \frac{P^+ + P^-}{|R|} \tag{14}$$

where P^+ and P^- refer to the number of correct predictions and incorrect predictions, respectively, and $|R|$ points out the total number of requests.

C. Evaluation of Next Location Prediction

The first experiment is designed to verify the performance of our model on location prediction. We use dataset D_3 for this part of experiment such that 70% of the data is used for the training, and the remaining 30% is for the prediction.

We compare our prediction model, "SPM + GSPM", with the traditional widely used approach which apply the general sequential prediction model (SPM), i.e. [5], [6], [7]. We consider this approach as the first baseline, base1. As to our own two-stage prediction framework, we also consider the related strategies proposed in this article as the additional baselines. The base2 considers only the group-based sequential prediction model (GSPM), and base3 considers both the general sequential prediction model and group-based sequential prediction model without applying the background smoothing. The third baseline also clarifies the importance of the background smoothing on performance of the prediction.

TABLE X
TABLE OF NOTATIONS

Parameter	Description	Default value
N	Number of groups	10000
$\|W\|$	$\|W\| \times \|W\|$ size of the network	15
$\|l\|, \|W\|$	Length and width of each cell	10, 10
U_I	Interestingness value	500
P_{Event}	Probability of group-based movement	0.7
W_b	The weight of backward movement	0.7
U_T	Length of Movement Path	20
N_{tp}	The number of time points in a day	200
N_T	Stay time in cell	20 minutes

TABLE XI
DEFAULT VALUES FOR THE PARAMETERS

Parameter	Description	Default value
$timeslot$	Duration of a time slot	1 minute
m_G	Group size threshold	2
d_G	Duration threshold	20 time slots
f_C	Frequency threshold	10 time slots
l_C	Gap threshold	10 time slots

TABLE XII
DEFAULT VALUES FOR THE PARAMETERS

Parameter	Description	Default value
δ	Support threshold	0.1%
st_U	Stay time	10 minuts

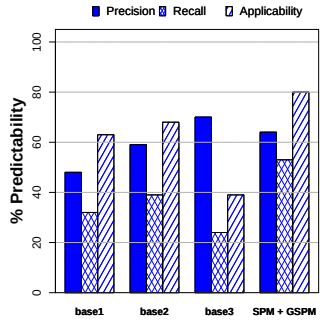


Fig. 8. Precision, Recall, and Applicability of different approaches

Fig. 8 represents the predictability of different models according to the default setting. As the figure indicates base1, base2, base3, and SPM + GSPM have achieved the precision of 48%, 59%, 70%, 64%, recall of 32%, 39%, 24%, 53%, and applicability of 63%, 68%, 39%, 80%, respectively. We observe that base3 achieves the highest precision, however, the recall and applicability of this approach are significantly lower than the other methods. The reason is, this approach predicts only the locations that can be found by both the sequential prediction model (SPM) and group-based prediction model (GSPM), which results in the precise prediction but low predictability. Our prediction model, instead, reaches the slightly lower precision compared to base3, while results in the best recall and applicability. Compared to base1 (widely used approach), SPM + GSPM improves the precision, recall, and applicability by 16%, 21%, and 17%, respectively.

First, we analyze the influence of database size on the performance of the prediction of different approaches. We change the size of the database between 1000 groups to 10000 groups and evaluate the precision, recall and applicability of four prediction methods. Fig. 9 shows that SPM + GSPM outperforms the baselines in terms of recall and applicability with varied database size. With increasing the database size, the precision is nearly constant. On the other hand, the recall and applicability of different approaches increases with increasing the database size. The reason is, with the availability of more movement data, the repeatability of the movement pattern increases which leads to the higher recall and applicability. Taking the precision into account, base3 achieves the higher result than our approach (the reason is the same as above), however, its recall and applicability are significantly lower than SPM + GSPM.

We also investigate the precision, recall and applicability when the event probability P_{Event} varies between 0.5 to 0.9. Fig. 10(a), Fig. 10(b), Fig. 10(c) show that predictability of our model outperforms the baselines in terms of recall and applicability with varied event probability. Considering the precision, however, base3 reveals the better results which explained above. We observe that precision, recall and applicability of the base2, base3 and "SPM + GSPM" increase with the varying the event probability from low values to the higher values. The reason is, when the event probability

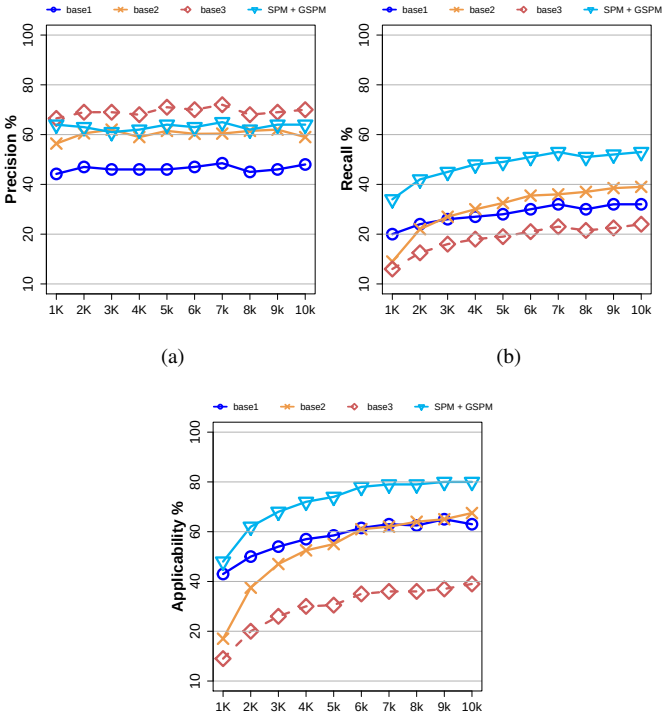


Fig. 9. Predictability: (a) precision, (b) recall, (c) applicability, vs DB size

increases, movement of the people is more affected by the type of the group they travel with which leads to the discovery of more precise group-specific prediction rules. Therefore the prediction models which are affected by the the group-specific prediction (base2, base3, and "SPM + GSPM") results in the higher precision, recall and applicability, with increasing the P_{Event}. On the other hand, the predictability of the general sequential prediction model (base1) degrades with increasing the event probability. The reason is the movement of people are less affected by only the sequential influence, which leads to the less comprehensive sequential prediction rules.

Finally, we perform the experiment to analyze the $precision$, $recall$, and $applicability$ of the prediction approaches with changing the size of the network, $|W|$. This parameter has the opposite effect of database size. Fig. 11(a), 11(b) and 11(c) reveal that our approach outperforms base1 and base2 in terms of precision, recall and applicability with varying the network size. Even though the precision of base3 is higher than SPM+GSPM (the reason is explained in subsection C), the recall and applicability of this approach are significantly lower than SPM+GSPM. While the precision is nearly constant with increasing the size of the network, both Recall and Applicability decrease with increasing the network size. The reason is, users in larger networks have different movement behaviours. Therefore the repeatability of the movement decreases which leads to the reduction in recall and applicability.

In conclusion, comparing to the base1, which is the tra-

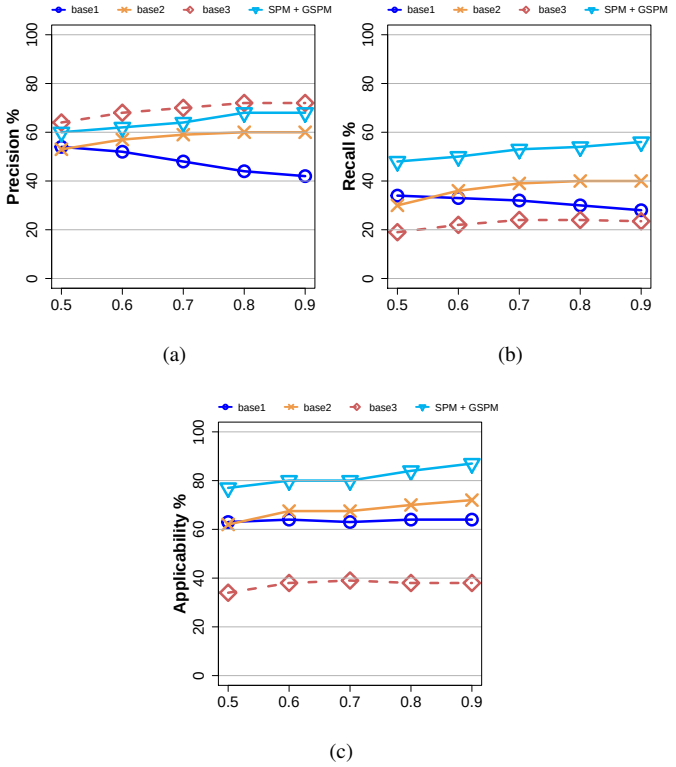


Fig. 10. Predictability: (a) precision, (b) recall, (c) applicability, vs P_{Event}

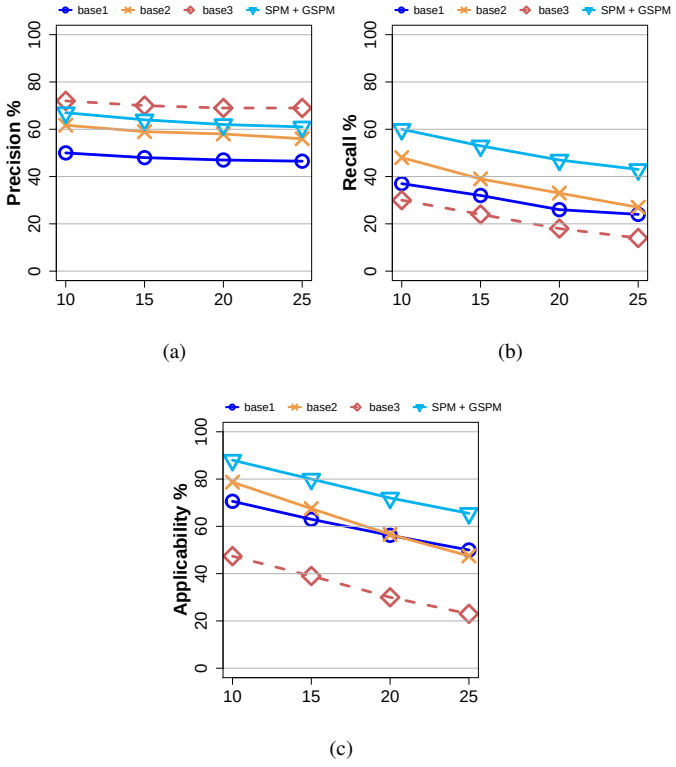


Fig. 11. Predictability: (a) precision, (b) recall, (c) applicability, vs $|W|$

ditional sequential based prediction, SPM+GSPM achieved significant improvement in terms of precision, recall and applicability. This superiority becomes clearer with increasing the value of P_{Event}.

D. Evaluation of Group Discovery

We compare the proposed group discovery method $WCLTCP$ with two state-of-the-art baselines: 1) the traveling companion pattern (TC) [44] which captures the groups whose members are close together for certain consecutive time intervals, 2) the loose companion pattern (LC) [49] which is the same as traveling companion pattern with the difference that the gatherings of the whole group can be non-strictly continuous, for certain time intervals. In this section, the quality of the discovered groups by different group discovery approaches will be evaluated. We consider the information of the groups in dataset D_1 as the ground truth, which the output of the different approaches will be compared with that. The following criterias will be used to evaluate the quality of the discovered groups, $Accuracy$ and $Precision$. We define the criterion of Accuracy as the proportion of the correct discoveries over the ground truth. We also measure the proportion of the correct discoveries over the retrieved results as $Precision$.

Fig. 12 plots the accuracy and precision of different group discovery approaches. As the figure indicates, TC, LTC, and and WCLTCP achieve the accuracy of 42%, 50%, 93% and the precision of 8.5%, 7%, and 70%, respectively. As the figure shows, WCLTCP significantly outperforms TC and LTC in terms of accuracy and precision. Because of the relaxed time constraint in LTC, it achieves better accuracy than the TC approach. However, it also leads to more false positive results, which results in the lower precision than TC. For all the approaches, the accuracy is considerably higher than the precision. It means that even though the ground truth groups are well discovered, group discovery approaches also result in wrong discoveries which leads to the lower precision. This difference is more severe for the baseline approaches such that although they are able to discover almost half of the ground truth, more than 90% of the discovered groups are wrong.

For further clarification, we show the accuracy and precision of the group discovery approaches in terms of different group sizes, Fig. 13. As the group discovery approaches extract groups with minimum size 2, we consider individuals who are not assigned to any groups as the groups with size 1 (solo traveler). There is a decreasing trend in terms of accuracy with increasing the group size, Fig. 13(a). It reveals that the group discovery approaches are less able in discovery of the larger groups than the smaller groups. However, WCLTCP shows a slight decrease, while TC and LTC reveal an abrupt reduction in accuracy with increasing the group size. In terms of precision, except the groups with size 1, WCLTCP achieves the precision nearly to 100%. Unlike the accuracy, precision shows the increasing trend as groups grow larger, Fig. 13(b). It reveals that, the wrong discovery is reduced with increasing the size of the group.

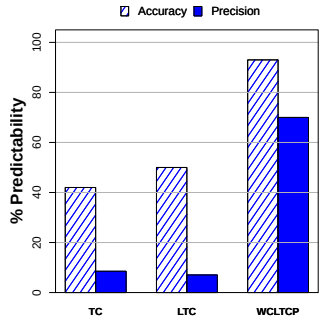


Fig. 12. Accuracy and Precision of different approaches

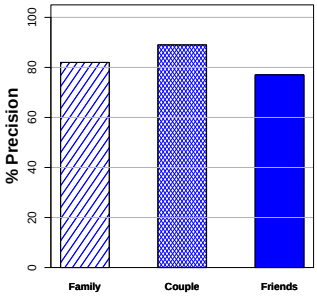
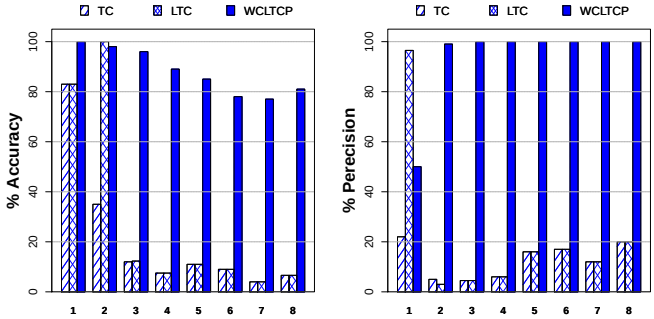


Fig. 14. Precision of predicting the group types



(a) (b)

Fig. 13. (a) accuracy, (b) precision vs group size

It should be noted, as our focus in this paper was on the prediction part, we omitted the detailed evaluation of the proposed group discovery. The comprehensive evaluation results can be found in [47].

E. Evaluation of Group Type Prediction

We evaluate the group type prediction performance by 10-fold cross validation over the three group types, family, friends, and couple on dataset D_2. As shown in Fig. 14, the accuracy for couple group is slightly better than the others, possibly due to the distinct features of the couple group, e.g., gender difference, which is a common characteristic in most couples. The prediction accuracy, on the average, can achieve 86%. It is clear that if more training data is available, the prediction accuracy can be further improved.

X. Conclusion

In this paper, we defined a new kind of frequent pattern, namely GFSP Pattern, which takes into account the group travel type of the users. Accordingly, we proposed a novel personalized prediction framework to predict the next location of a user for applications such as location-based services. The core idea of our prediction module is a novel prediction strategy that evaluates the score of the next location for a given user by mining the movement patterns of users in terms of the general and group-specific properties. To discover the travel groups from spatial-temporal trajectories, we applied a novel group pattern discovery which takes the human movement behaviour into consideration. Then, according to the profile information of the individual, the type of the discovered groups was identified. To the best of our knowledge, this is the first work that focuses on next location prediction by mining trajectory data that takes into consideration the groups users travel with. We evaluated our prediction model through a series of experiments and showed that our approach outperforms the widely used sequential approach in terms of precision, recall, and applicability. For the future work, we want to expand our prediction model with more contextual data such as the travel duration. We believe such location prediction models which are enriched with the contextual information are promising for the location-based applications such as advertisement. Besides, since our prediction model further encompasses group discovery from the movement trajectories, a number of parameters corresponding to that are used in the model. As the next step, we will also try to reduce the complexity of the model by simplifying the parameters.

ACKNOWLEDGEMENTS

The authors would like to thank Wuxi Chigoo Interactive Technology Co. Ltd to sponsor a research project to conduct research in this area, provide data for analysis and also implement and test the algorithms. The authors would also like to thank Royal Society of Edinburgh and National Natural Science Foundation of China under grant number 6151101271 to engage academic staff from the UK and China to work together on this project. In addition, acknowledgements also go to National Natural Science foundation of China to sponsor the research in part under grant numbers 6177020565, U1509215, and 61401004.

References

[1] W. C. Peng and M. S. Chen, "Developing data allocation schemes by incremental mining of user moving patterns in a mobile computing system", *IEEE Trans. on Knowledge and Data Engineering*, vol. 15, no. 1, pp. 70-85, Feb. 2003.

[2] Y. GE, H. XIONG, A. TUZHILIN, K. XIAO, M. GRUTESER, and M. J. PAZZANI, " An Energy-Efficient Mobile Recommender System," *In proceeding of KDD*, 2010.

[3] Q. LIU, Y. GE, Z. LI, E. CHEN, and H. XIONG, " Personalized Travel Package Recommendation," *In proceeding of ICDM*, 2011.

[4] J. ZHUANG, T. MEI, S. HOI, Y. XU, and S. LI, " When recommendation meets mobile: Contextual and personalized recommendation on the go," *In Proceedings of UbiComp*, 2011.
[5] G. Yavas, D. Katsaros, O. Ulusoy, and Y. Manolopoulos, "A data mining approach for location prediction in mobile environments", *Data and Knowledge Engineering*, vol. 54, no. 2, pp. 121-146, 2005.
[6] M. Morzy, "Prediction of moving object location based on frequent trajectories", *ISCIS* , vol. 4263 of LNCS, pp. 583-592, 2006.
[7] M. Morzy, "Mining frequent trajectories of moving objects for location prediction", *MLDM* , vol. 4571, pp. 667-680, 2007.
[8] JJ. Ying, WC. Lee, and VS. Tseng, "Mining geographic-temporal-semantic patterns in trajectories for location prediction,"*ACM Transactions on Intelligent Systems and Technology*, 2013.
[9] A.J. Cheng, Y.Y. Chen, Y.T. Huang, W.H. Hsu, and H.Y. M. Liao, "Personalized travel recommendation by mining people attributes from community-contributed photos,"*In ACM*, 2011.
[10] Y.Y. Chen, A.J. Cheng, and W.H. Hsu, "Travel recommendation by mining people attributes and travel group types from community-contributed photos,"*IEEE Transactions on Multimedia*, pp.1283-1295, 2013.
[11] J. M. Patel, Y. Chen, and V. P. Chakka, "Stripes: an efficient index for predicted trajectories,"*in SIGMOD*, pp. 635-646, 2004.
[12] C. S. Jensen, D. Lin, and B. C. Ooi, "Query and update efficient B+-tree based indexing of moving objects", *in VLDB*, pp. 768-779, 2004,
[13] S. Saltenis, C. S. Jensen, S. T. Leutenegger, and M. A. Lopez, "Indexing the positions of continuously moving objects,"*in SIGMOD*, 2000, pp. 331-342.
[14] Y. Tao, D. Papadias, and J. Sun, "The tpr*-tree: An optimized spatio-temporal access method for predictive queries,"*in VLDB*, pp. 790-801, 2003.
[15] S.M. Ghoreyshi, A. Shahrabi, and T. Boutaleb, "An Underwater Routing Protocol with Void Detection and Bypassing Capability", *In IEEE 31st International Conference on Advanced Information Networking and Applications (AINA)*, pp. 530-537, 2017.
[16] Y. Tao, C. Faloutsos, D. Papadias, and B. Liu, "Prediction and indexing of moving objects with unknown motion patterns,"*in SIGMOD*, pp. 611-622, 2004.
[17] C. C. Aggarwal and D. Agrawal, "On nearest neighbor indexing of nonlinear trajectories,"*in PODS*, pp. 252-259, 2003.
[18] C. Song, Z. Qu, N. Blumm, and A.L. Barab'asi, "Limits of Predictability in Human Mobility,"*Science 327*, no. 5968, pp. 1018-1021, 2010.
[19] A. Asahara, A. Sato, K. Maruyama, and K Seto, "Pedestrian-movement Prediction based on Mixed Markov-chain Model,"*In Proc. of ACM-GIS*, pp. 25-33, 2011.
[20] D. Ashbrook and T. Starner, "Using GPS to Learn Significant Locations and Predict Movement across Multiple Users,"*Personal and Ubiquitous Computing* , pp. 275-286, 2003.
[21] A. Bhattacharya and S. K. Das, "LeZi-Update: An Information-Theoretic Approach to Track Mobile Users in PCS Networks,"*In Proc. of MobiCom*, pp. 1-12, 1999.
[22] Y. Ishikawa, Y. Tsukamoto, and H. Kitagawa, "Extracting Mobility Statistics from Indexed Spatio-Temporal Dataset,"*In Proc. of STDBM*, pp. 9-16. 2004.
[23] F. Giannotti, M. Nanni, F. Pinelli, AND D. Pedreschi, " Trajectory pattern mining"*In Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 330-339, 2007.
[24] F. Verhein, and S. Chawla, "Mining Spatio-Temporal Association Rules, Sources, Sinks, Stationary Regions and Thoroughfares in Object Mobility Databases"*In Proc. of DASFAA*, pp. 187-201, 2006.
[25] Y. Zheng, L. Zhang, X. Xie, and W.Y. Ma, "Mining Interesting Locations and Travel Sequences from GPS Trajectories"*In Proceedings of the 18th international conference on World wide web*, pp. 791-801, 2009.
[26] Y. Ye, Y. Zheng, Y. Chen, J. Feng and X. Xie, "Mining Individual Life Pattern Based on Location History"*In Mobile Data Management: Systems, Services and Middleware,*, pp. 1-10, 2009.
[27] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Baraba'si, "Understanding individual human mobility patterns"*Nature*, pp. 779-782, 2008.
[28] H. Cao, N. Mamoulis, and D. W. Cheung, "Mining frequent spatio-temporal sequential patterns", *In Fifth IEEE International Conference on Data Mining*, 2005.
[29] E. Naserian, X. Wang, X. Xu, Y. Dong, N. Georgalas, and K. Huang, "Integrated Discovery of Location Prediction Rules in Mobile Environment", *In Third IEEE International Conference on Big Data Intelligence and Computing and Cyber Science and Technology*, pp. 1017-1024, 2017.
[30] R. Agrawal, R. Srikant, "Mining sequential patterns," *in Proceedings of the IEEE Conference on Data Engineering (ICDE95)*, pp. 3-14, 1995.
[31] J. Pei, J.Han, B. Mortazavi-asl, H. Pinto, Q. Chen, U. Dayal, and M.C. Hsu, "PrefixSpan: Mining sequential patterns efficiently by prefix-projected pattern growth", *In Proceedings of the 17th International Conference on Data Engineering* , pp. 215-224, 2001.
[32] H. Jeung, Q. Liu, H. T. Shen, and X. Zhou, "A hybrid prediction model for moving objects", *ICDE*, pp. 70-79, 2008.
[33] Y. Wu, G. Min, K. Li, and B. Javadi, "Modeling and Analysis of Communication Networks in Multicluster Systems under Spatio-Temporal Bursty Traffic,"*IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 5, pp. 902-912, 2012.
[34] F. GiannottiI, M. Nanni, and D. Pedreschi, "Efficient mining of temporally annotated sequences", *In Proceedings of the 6th SIAM International Conference on Data Mining*, 2006.
[35] J. J.C. Ying, W.C. Lee, T.C. Weng. and V. S. Tseng, "Semantic Trajectory Mining for Location Prediction", *In Proc. of ACM-GIS*, pp. 34-43, 2011.
[36] V. S. Tseng, and C. F. Tsui, "Mining Multi-Level and Location-Aware Associated Service Patterns in Mobile Environments", *IEEE Trans. on Systems, Man and Cybernetics: Part B*, vol. 34, no. 6, 2004.
[37] V. S. Tseng and W. C. Lin, "Mining Sequential Mobile Access Patterns Efficiently in Mobile Web Systems", *Proc. 19th Int. Conf. on Advanced Information Networking and Applications*, pp. 867-871, 2005.
[38] V. S. Tseng and K. W. Lin, "Efficient Mining and Prediction of User behaviour Patterns in Mobile Web Systems", *Information and Software Technology*, vol. 48, no. 6, pp. 357-369, 2006.
[39] J. Han, J. Pei, and Y. Yin, "Mining Frequent Patterns without Candidate Generation", *Proc. ACM SIGMOD Conf. on Management of Data*, pp. 1-12, 2000.
[40] S. C. Lee, J. Paik, J. Ok, I. Song, and U. M. Kim, "Efficient Mining of User behaviours by Temporal Mobile Access Patterns", *Int. J. of Computer Science Security*, vol. 7, no. 2, pp. 285-291, 2007.
[41] C. H. Yun and M. S. Chen, "Mining Mobile Sequential Patterns in a Mobile Commerce Environment", *IEEE Trans. on Systems, Man, and Cybernetics*, Part C, vol. 37, no. 2, pp. 278-295, 2007.
[42] E. Lu, V. Tseng and P. Yu, "Mining cluster-based temporal mobile sequential patterns in location-based service environments", *IEEE Transactions on Knowledge and Data Engineering*, pp. 914-927, 2011.
[43] L. Backstorm, E. Sun, and C. Marlow, "Find me if you can: Improving geographical prediction with social and spatial proximity", *In Proceedings of the 19th International Conference on World Wide Web*, 2010.
[44] Monreale, Anna, et al, "Wherenext: a location predictor on trajectory pattern mining", *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining,*, 2009.
[45] H. Jeung, M. Yiu, X. Zhou, C. Jensen, and H. Shen, "Discovery of convoys in trajectory databases," *Proceedings of the VLDB Endowment*, vol. 1, no. 1, pp. 1068-1080, 2008.
[46] L. A. Tang, Y. Zheng, J. Yuan, J. Han, A. Leung, C.C. Hung, and W. C. Peng, "On discovery of traveling companions from streaming trajectories," *In 28th IEEE International Conference on Data Engineering*, pp. 186-197, 2012.
[47] E. Naserian, X. Wang, X. Xu, Y. Dong, "Discovery of Loose Travelling Companion Patterns from Human Trajectories", *In 14th IEEE International Conference on Smart City*, 2016.
[48] X. Wang, C. Zhang, F. Liu, Y. Dong, and and X. Xu, "Exponentially Weighted Particle Filter for Simultaneous Localization and Mapping Based on Magnetic Field Measurements," *IEEE Transactions on Instrumentation and Measurement*, 2016.
[49] L. A. Tang, Y. Zheng, J. Yuan, J. Han, A. Leung, W. C. Peng, and T. L. Porta, "A framework of traveling companion discovery on trajectory data streams," *ACM Transactions on Intelligent Systems and Technology (TIST)* , vol. 5, no. 1, p. 3, 2013.

Elahe Naserian is currently working toward the Ph.D. degree from the School of Engineering and Computing, University of the West of Scotland, United Kingdom. She obtained her master's degree from the University of Tehran, Iran.

Her research interests include data mining and knowledge discovery over spatial-temporal data.

Xinheng Wang (M'04-SM'14) received the B.E. and M.Sc. degrees in electrical engineering from Xian Jiaotong University, Xian, China, in 1991 and 1994, respectively, and the Ph.D. degree in electronics and computer engineering from Brunel University London, Uxbridge, U.K., in 2001. He is currently a Professor in Computing with the School of Computing and Engineering, University of West London, London, U.K. He holds seven patents and has authored or co-authored over 150 referred papers. He has broad research experience in mobile healthcare, asset monitoring, and wireless mesh and sensor networks, where the technologies developed in wireless mesh networks have been commercialized at Swanmesh Networks Ltd (www.swanmesh.com). His current research interests include indoor positioning, Internet-of-Things, and big data analytics for smart airport services, where he has developed the world's first smart trolley with industry partner.

Keshav Dahal is a Professor of Intelligent Systems and the leader of the Artificial Intelligence, Visual Communication and Network (AVCN) Research Centre at the University of the West of Scotland (UWS), UK. He is also affiliated with Nanjing University of Information Science and Technology (NUIST) China. Before joining UWS he was with Bradford and Strathclyde Universities in UK. He obtained his Ph.D. and Master degrees from the University of Strathclyde, UK. His research interests lie in the areas of applied AI to intelligent systems, trust and security modelling in distributed systems, and scheduling/optimization problems. He has published extensively with award winning papers, and has sat on organizing/program committees of over 60 international conferences including as the General Chair and Programme Chair. He is a senior member of the IEEE.

Zhi Wang is currently an associate professor at State Key Laboratory of Industrial Control, Zhejiang University, China. His main research focus is on acoustic signal and array processing, sparsity signal and compressive sensing, localization and tracking of mobile target, multiple sensor fusion, crowd-sensing and mobile computing, big-data and industrial IoT protocols. He has co-authored over 100 publications in international journals and conferences and also has served as an member of advisory board and an editor of several conferences. He is also the committee member for China Computer Federation Sensor Network Technical Committee and China National Technical Committee of Sensor Network Standardization, and is a member of IEEE and ACM.

Zaijian Wang received his BE degree (2002) from Anhui Polytechnic University, MSc degree (2005) from University of Science and Technology of China, and PhD degree (2015) from Nanjing University of Posts and Telecommunications. His current research interests focus on multimedia big data, end-to-end QoS provisioning and wired/wireless multimedia streaming. He is currently an associate professor at College of Physics and Electronic Information, An'hui Normal University, Wuhu, China.