

## Location Prediction on Trajectory Data: A Review

Ruizhi Wu, Guangchun Luo\*, Junming Shao, Ling Tian, and Chengzong Peng

**Abstract:** Location prediction is the key technique in many location based services including route navigation, dining location recommendations, and traffic planning and control, to mention a few. This survey provides a comprehensive overview of location prediction, including basic definitions and concepts, algorithms, and applications. First, we introduce the types of trajectory data and related basic concepts. Then, we review existing location-prediction methods, ranging from temporal-pattern-based prediction to spatiotemporal-pattern-based prediction. We also discuss and analyze the advantages and disadvantages of these algorithms and briefly summarize current applications of location prediction in diverse fields. Finally, we identify the potential challenges and future research directions in location prediction.

**Key words:** location prediction; trajectory data; data mining

### 1 Introduction

Urban planning, relieving traffic congestion, and effective location recommendation systems are important objectives worldwide and have received increasing attention in recent years. Spatiotemporal data mining is the key technique involved in these practical applications<sup>[1–3]</sup>. Trajectory data brings new opportunities and challenges in the mining of knowledge about moving objects. To present, many researchers have used trajectory data to mine latent patterns that are hidden in data. These patterns can also be extracted for the analysis of the behavior of moving objects. Location prediction, as the primary task of spatiotemporal data mining, predicts the next location of an object at a given time. In recent years, researchers in location prediction have

made much progress. For instance, early studies traced student ID cards to identify frequent temporal patterns and used these patterns to predict their next location<sup>[4–7]</sup>. Since then, location prediction has had a wide range of applications in daily life, e.g., travel recommendation, location-aware advertisements, and early warning of potential public emergencies, to mention a few<sup>[8–10]</sup>. Location prediction typically must employ many techniques, including trajectory data preprocessing, trajectory clustering, trajectory pattern mining, trajectory segmentation, and trajectory semantic representation. In this article, we review the field of location prediction, its basic definitions, typical algorithms, model evaluations, and diverse applications. Our objective in this review is to present a comprehensive picture of location prediction.

The remainder of this paper is organized as follows: in Section 2, we introduce the basic concepts of location prediction, including the different sources of trajectory data, the general prediction framework, challenges in location prediction, and common trajectory data preprocessing methods. In Section 3, we introduce some common trajectory data preprocessing methods. Section 4 comprises the core of this review, in which we describe different models and briefly introduce each of their motivation, basic idea, and key techniques. In

• Ruizhi Wu, Guangchun Luo, Junming Shao, Ling Tian, and Chengzong Peng are with School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China. E-mail: ruiwzhi@uestc.edu.cn; gcluo@uestc.edu.cn; gcluo.uestc@gmail.com; junmshao@uestc.edu.cn; lingtian@uestc.edu.cn; prescott0307@outlook.com.

\* To whom correspondence should be addressed.

Manuscript received: 2017-11-20; accepted: 2018-01-10

Section 5, we provide some public data sets, and typical evaluation strategies used in location prediction. Finally, in Section 6, we introduce some real-world applications and future directions. We conclude our overview in Section 7.

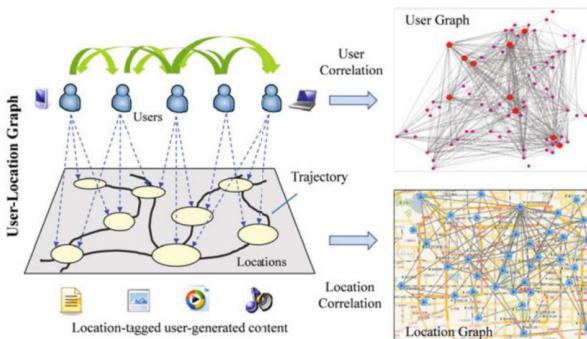
## 2 Preliminary

Trajectory data characterizes the locations and times of moving objects. Formally,  $p = (x, y, h, t)$  represents a trajectory data point, where  $x$  and  $y$  are the latitude and longitude of a given moving object, respectively.  $h$  is the altitude, and  $t$  is the time stamp (in many real-world applications,  $h$  is ignored; thus  $p = (x, y, t)$  is commonly used). As such, trajectory data is composed of a sequence of trajectory data points in chronological order. Formally, a trajectory  $T_{ra}$  is often represented as:  $T_{ra} = p_1, p_2, \dots, p_i, \dots, p_n$ . In the following subsection, we introduce different types of trajectory data sources, their unique characteristics, and the corresponding challenges in location prediction.

### 2.1 Trajectory data sources

There are many different types of trajectory data in real-world scenarios. Zheng<sup>[11]</sup> classified trajectory data as being either actively or passively recorded, depending on the derivation of trajectories. We briefly introduce these two types of trajectory data below.

- *Active recording trajectory data:* People actively record their locations when they login to social networks or travel to places of interest and share their life experiences. Typical data types include check-in data (e.g., Twitter, Weibo, QQ, etc.), and location-based data such as travel photos. Figure 1 shows the correlation of users and locations in social networks. In Flickr, a sufficient number of geotagged photos can be formulated



**Fig. 1** User-location graph for a Location-Based Social Network (LBSN), comprising user and location correlations<sup>[14]</sup>.

as a trajectory, whereby each photo is associated with a location tag and a time stamp. Due to the random behaviors of users, active recording data is typically characterized by its sparsity. When mining this trajectory data, additional social information is usually added.

- *Passive recording trajectory data:* With the development of positioning techniques, many moving objects are equipped with positioning position devices that record location information. These include global positioning systems GPS in vehicles and radio-frequency identification devices for tracing objects. Typically, these devices automatically record huge volumes of trajectory data points<sup>[12, 13]</sup>. Transaction records or the Internet trajectory of human beings can also be viewed as trajectory data, since locations and times are recorded.

### 2.2 Challenges in location prediction

In contrast to traditional data, the unique properties of trajectory data (e.g., different sampling rates, different lengths, and sparsity) mean that location prediction faces many challenges.

#### 2.2.1 Randomness of movement behaviour

Since the current location of a user is related to that users visited location history, trajectory data is context-sensitive. However, in contrast to other data, mobility patterns are difficult to identify. Features in trajectory data are fuzzy, and mobility patterns have no exact standards. To generate patterns, researchers often mine mobility patterns based on association rules, time period, and transition probability between locations. Sometimes patterns can represent user movement behaviour, but cannot describe the moving process of users due to the inherent mobility randomness of mobility.

#### 2.2.2 Time sensitivity

Trajectory data is time sensitive. Since the moving speed of a user can be fast, a user often visits many locations in a short period of time. The time sensitivity property makes it difficult to handle time. Traditional methods often use a time window, but this strategy is not always optimal, because (1) users move quickly and randomly, and a time window cannot capture changes in the trajectory data; (2) the length of a time window is difficult to establish; and (3) a time window is a discrete way to handle time and it is difficult to determine the

correlation between time and location.

### 2.2.3 Cold start and sparsity problems

If a user has no trajectory history, it is difficult to establish predictors regarding his future location, which is known as the cold start problem. If a user has only a few visiting locations, it is also difficult to establish predictors of his future location, which is a sparsity problem. Cold start and sparsity problems are prevalent in prediction applications, especially those using actively recorded trajectory data.

### 2.2.4 Heterogeneous data

The sources of trajectory data are diverse, e.g., taxis, buses, people, etc. These moving objects often have different sampling rates and movement patterns. In many scenarios, datasets also include social relations or short messages. As such, heterogeneous data sources represent another challenge in location prediction systems.

## 2.3 Location prediction framework

Human trajectory data reveal the movement preferences and behaviours of people in daily life. Trajectory data mining involves trajectory data processing, management, and pattern mining from past trajectories<sup>[11]</sup>. Location prediction, as the primary task in trajectory data mining, learns the movement patterns of moving objects based on their past locations and the past and then forecasts future locations. In traditional trajectory mining tasks, trajectory pattern mining, trajectory similarity measurement, and trajectory anomaly detection are closely related to location prediction. The objective of location prediction is to predict the next location and/or the next visit time to a given location. The former considers trajectory data as a spatial sequence, whereas the latter regards trajectory data as a spatiotemporal sequence.

Figure 2 shows the framework of the location prediction process, which has three steps. (1) Due to trajectory data being sampled from various positioning devices, its quality is low, so data preprocessing is necessary. Trajectory data preprocessing involves noise filtering, data cleaning<sup>[15]</sup>, trajectory data compression<sup>[16-18]</sup>, even some special preprocessing such as trajectory segmentation<sup>[19-21]</sup>, trajectory semantics<sup>[13, 22, 23]</sup> or map matching<sup>[24-26]</sup>. (2) Learning from past trajectory data is a modeling process, in which the movement of a moving object is modeled. The performance of a location prediction algorithm mainly depends on the proposed model. (3) Location prediction and results evaluation are the indispensable parts of an established location prediction system.

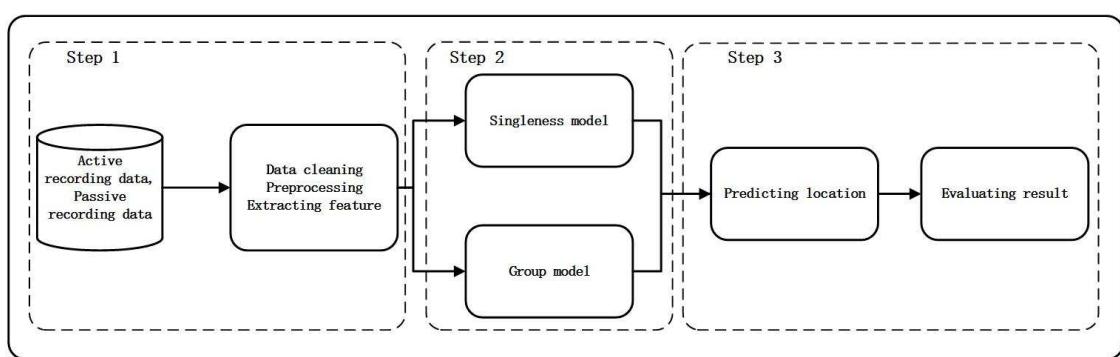
In this section, we review some basic concepts of location prediction, including the data sources and challenges associated with location prediction. Finally, we provided an overview of the location prediction framework.

## 3 Trajectory Data Preprocessing

Based on the above general framework, we briefly introduce some common trajectory data preprocessing methods in this section. First, we present some common basic data cleaning methods, including noise filtering, stay point detection, and trajectory compression. Then, we introduce some special trajectory data preprocessing procedures, which typically involve Places Of Interest (POIs) identification, trajectory segmentation, and trajectory semantics. Finally, we introduce methods for extracting features from trajectory data.

### 3.1 Data cleaning

Traditional trajectory data preprocessing often involves



**Fig. 2 Popular general framework of location prediction in which the integration of single object and group models is an emerging direction.**

noise filtering, stay point detection, and trajectory compression.

Noise filtering addresses data inaccuracy problems when the sampled trajectory data from positioning equipment contains unexpected mistakes, such as sensor noise. Existing methods are mainly categorized into three types: mean (or median) filters, Kalman filters, and particle filters.

The stay point is a trajectory point in which the geographical position does not change over a relatively long period of time. Generally, stay points usually have a special meaning to the moving objects, such as a work place, restaurant, or nest for animals. The earliest stay point detection algorithm was proposed by Li et al.<sup>[27]</sup> and Zheng<sup>[28]</sup>. Many stay point detection algorithms are based on the concepts of density and nearest neighbours.

To present, the volume of trajectory data is rapidly increasing due to the popularity of positioning devices. However, a lot of trajectory data suffers from high costs of storage and analysis. Trajectory data compression is used to extract key information from trajectory data, including some key points in a trajectory to briefly represent the trajectory. Existing methods can be divided into three categories: off-line, on-line, and semantic compression.

Table 1 summarizes some of the main data cleaning methods. For details, the reader is referred to the corresponding references.

### 3.2 Preprocessing for location prediction

After data cleaning, data preprocessing is usually necessary in location prediction. For example, the goal in most applications is to determine the next location at which people will arrive, rather than the specific GPS coordinates. In this scenario, trajectory preprocessing mainly includes map matching, POIs identification,

**Table 1** Summary of existing trajectory data cleaning methods.

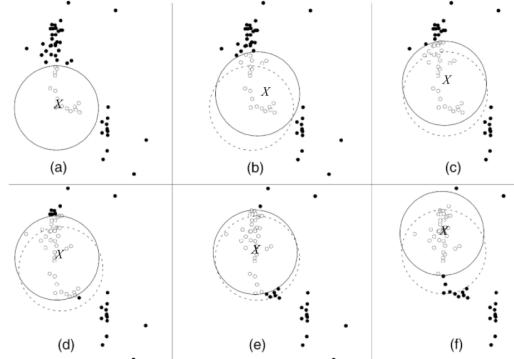
Data cleaning	Method
Noise filtering	Mean (or median) filters <sup>[11]</sup> Kalman filters <sup>[15, 29]</sup> Particle filters <sup>[30, 31]</sup> Density clustering <sup>[32]</sup>
Stay point detection	Nearest neighbour <sup>[33, 34]</sup> Content-based <sup>[35]</sup> Off-line compression <sup>[36]</sup> On-line compression <sup>[17, 37]</sup> Semantic compression <sup>[13, 28]</sup>
Trajectory compression	

trajectory segmentation, and trajectory semantics.

Urban trajectory data can be used to map a road network<sup>[38]</sup>. As such, it is beneficial to analyze the paths taken by people, and this also provides a new representation of the trajectory data, i.e., a graph. Map matching techniques map trajectory data onto a road network. Current mainstream matching algorithms are either context-based<sup>[39, 40]</sup> or integrate additional information<sup>[41, 42]</sup>.

POIs are great concern in urban places, and are also referred to as significant locations. POIs denote locations frequently visited or high-density trajectories near locations. Currently, POIs data can be divided into two types. The first type includes well-known public places like malls, government offices, and transportation centers, to mention a few. Unfortunately, this type of data is often scarce. The second type of POI data is high-density trajectory data, in which POIs are often invisible and a cluster-based algorithm or hierarchical method must be utilized to identify the POIs. Figure 3 outlines the process of identifying significant locations (POIs) via hierarchical clustering methods<sup>[43]</sup>, where  $X$  is the center of the cluster, and the dots indicate GPS data. In the figure, the white dots are the GPS data in a cluster, and the dots within the dotted line represent data in the previous cluster. The basic idea of clustering algorithms is to mark a place and its radius. GPS data points within this radius are recognized as the same POI, and the central point, which is new, represents the mean of these GPS data points. This process is repeated until all points within this radius no longer change. We note that POIs are different from stay points. A stay point is the cluster point of a trajectory, whereas POIs are key places of interest.

In some scenarios, since the local trajectory is



**Fig. 3** Process for identifying significant locations using cluster methods<sup>[43]</sup>.

more important than the whole trajectory, trajectory segmentation is introduced. Trajectory segmentation methods involve finding the key point, and minimizing the specific cost function or rule-based methods. Early studies identified an inflection point or angle-change distinct point as the key point. Since then, more mature techniques have been proposed. For example, Lee et al.<sup>[19]</sup> proposed a trajectory segmentation approach using MDL (Minimum Description Length) to minimize the cost of coding this trajectory. Trajectory segmentation is essential in local trajectory trend analysis and local trajectory clustering.

Trajectory data are closely related to the behavior of moving objects, especially humans<sup>[13, 44, 45]</sup>. Trajectory semantics identifies semantic information about object movement, to explain human behavior, identify the semantic pattern, and establish an outlook for exploring trajectory data. A number of researchers have enriched semantic information into trajectory via data mining. Trajectory semantics facilitates location prediction by its ability to interpret human behaviour. For more details, the reader is referred to the review of trajectory semantics<sup>[46]</sup>.

In this subsection, we introduced some new trajectory representation techniques, specifically, map matching of trajectories as a graph, the representation of trajectories as sequential POIs, trajectory segmentation by splitting trajectories into segments in a set, and trajectory semantics that consider a trajectory in a

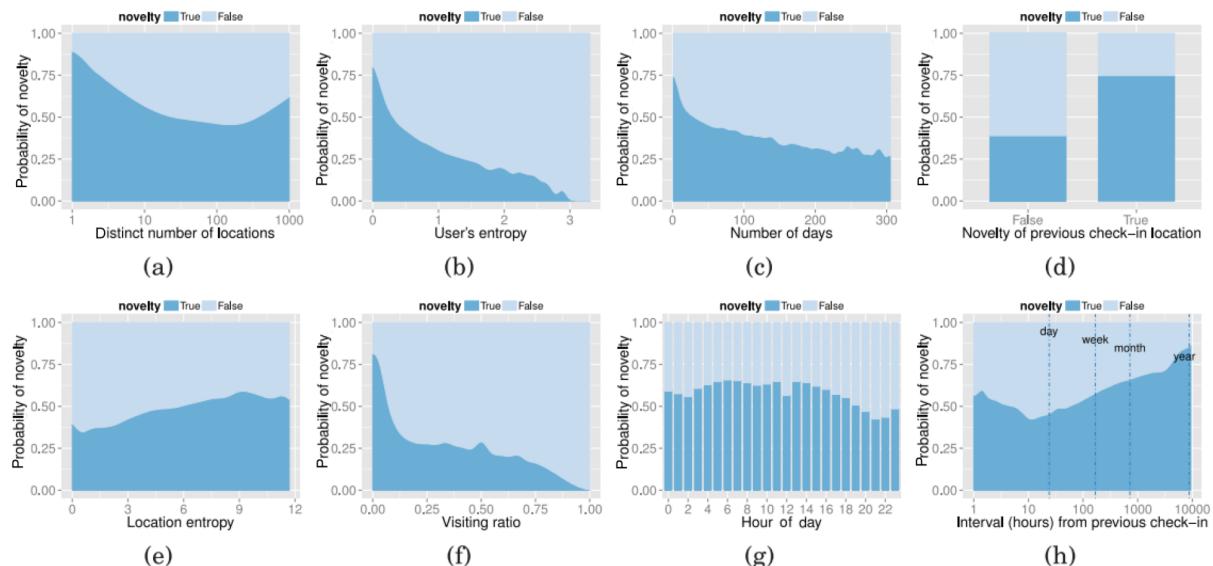
semantic representation.

### 3.3 Trajectory data features

In contrast to other data, it is a non-trivial task to extract features from trajectory data. Some researches define a start point, end point, moving speed, and time length as features. Wang et al.<sup>[47]</sup> proposed new features in trajectory data, consisting of spatial mobility patterns, text content, individual temporal patterns, social relationships, collaborative filtering, and heterogeneous mobility datasets. Social relationships describe the social networks in Location Based Social Network (LBSN), in which the extracted features include the number of locations, user entropy, location entropy, and visitation ratio. Figure 4 shows some features in the Gowalla and Jiepang datasets<sup>[48]</sup>. There are various types of heterogeneous mobility datasets including bus, taxi, check-in, and life mobility datasets. Spatial mobility patterns, individual temporal patterns, and text content are extracted from individual trajectories via data mining or machine learning and collaborative filtering techniques are used to find similar users or locations. These features provide unique insights for exploring trajectories.

## 4 Location Prediction Models

In this section, we provide a summary of current location prediction algorithms, introduce some representative studies, and discuss the categorization of



**Fig. 4** Influence of different features on novel check-in location behaviours, x-axes are features and y-axes are the conditional probabilities. (a) Distinct number of locations, (b) user entropy, (c) number of days, (d) novelty of check-in location, (e) location entropy, (f) visiting ratio, (g) hour of day, and (h) interval since previous check-in<sup>[48]</sup>.

prediction models.

#### 4.1 Location prediction algorithms

Here, we introduce some representative location prediction algorithms. Table 2 lists the main symbols used.

##### 4.1.1 Content-based methods

Content-based methods learn the content correlation or location transition probability, based on the assumption that the current user location is related to the previous location. To this end, researchers often construct a data structure to store content, then match the content to predict location<sup>[49,50]</sup>. The Markov model is a typical strategy that uses content information to predict a future state.

Song et al.<sup>[43]</sup> studied mobile phone positioning data via Wi-Fi signal, using a two-year trajectory trace of more than 6000 users on Dartmouth College's campus-wide Wi-Fi wireless network. The authors compared two prediction methods, Markov-based and compression-based location predictors. A Markov model is a typical sequence analysis method in which a Markov chain is used to model a user's location history based on the user's transition probability from one location to another. In this model, it is assumed that a user's next location depends on his current location. If a user arrives at an associated location, he is likely to visit the identified location. A  $k$ -order Markov model enables the prediction of a user location based on previous  $k-1$  locations. If a user is currently at the  $n$ -th location, his  $k$ -order recent locations are  $(l_{n-k+1}, \dots, l_n)$ . Formally, if we assume that a user trajectory history is  $l_1, l_2, l_3, \dots, l_{n-k+1}, \dots, l_n$ , the  $k$ -order context is  $c = (l_{n-k+1}, \dots, l_n)$  and  $L$  is the location sets, then the Markov model is defined as shown in Eq. (1):

$$\begin{aligned} P(l_{n+1} = l^* | L(1, n)) &= \\ P(l_{n+1} = l^* | c = (l_{n-k+1}, l_n)) &= \\ P(l_{i+k+1} = l^* | c = (l_{i+1}, l_{i+k})) & \end{aligned} \quad (1)$$

**Table 2 Main symbols used in location prediction models.**

Symbol	Description
$P$	Probability
$l$	Location
$L$	Location sets
$c$	Content of location
$Tra$	Trajectory
$st$	Sub-trajectory
$h$	History trajectory
$u$	User

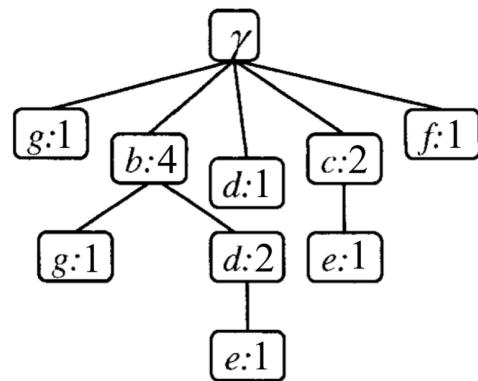
where  $P(l_{n+1} = l^* | \dots)$  is the probability of the user arriving at the next place  $l^*$ . The first line in Eq. (1) is the Markov assumption, and the second line indicates that there is the same probability of arriving at a different location if the context is similar. Moreover, if a location has no historical context, its arrival probability is equal to zero.

The other content-based method uses compression-based location predictors with a popular incremental parsing algorithm for text compression. In this method, a trajectory is first partitioned into distinct sub-trajectories  $st_0, st_1, \dots, st_m$ . Let  $l_0 = \gamma$ , and for  $j > 0$ , sub-trajectory  $l_j$  removes its last location that is the same as  $l_i$ , where  $0 < i < j$  and  $st_1st_2\dots st_m = Tra$ . For example, for a trajectory  $gbdcbgcefbdedbde$ , after partitioning by the above rule, the sub-trajectory sets are  $g, b, d, c, bg, ce, f, bd, bde$ . Then, a tree is built to store these sub-trajectories, in which the root nodes are  $\gamma$ , and the child node is sub-trajectories and each with a location occurrence number. Figure 5 shows an example.

For each location  $l^*$  in the location set, Eq. (2) is used to compute the occurrence probability, where  $N(st_m, l^*)$  denotes the number of  $st_m$  occurring as a prefix for  $st_m, l^*$  in the sub-trajectory data sets.

$$P(l_{n+1} = l^* | L) = \frac{N(st_m, l^* | L)}{N(st_m | L)} \quad (2)$$

Unlike Markov-based predictors which have a fixed length related to the  $k$  location, compression-based predictors depend on the occurrence number of location prefixes. Song et al.<sup>[43]</sup> compared the performance of these two methods in predicting the location of users via user Wi-Fi data. Based on their experimental results, the authors found low-order Markov methods to be more accurate than complex compression-based methods. Next, research is conducted to determine the

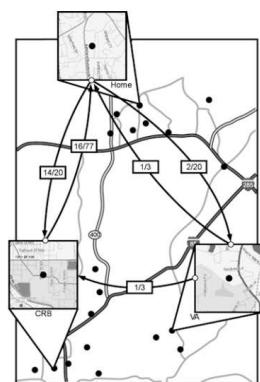


**Fig. 5 Example of a tree storage sub-trajectory based on compression<sup>[43]</sup>.**

transition probability between different locations. The transition probability is used in both Markov-based and compression-based predictors, and is a prototype of the Spatial Mobility Pattern (SMP) feature that provides a new way of investigating changes in movement patterns over time.

Similar to the findings of Song et al.<sup>[43]</sup>, Ashbrook and Starner<sup>[51]</sup> proposed a location prediction model based on a Markov chain, which they called the user location pattern ULP model, and which has wide application in many scenarios. In this model GPS data is transmitted to significant locations. These locations are then input to a Markov model to predict locations. The difference between this model and that proposed by Song et al.<sup>[43]</sup> is that ULPs are applied to both single and collaborative users. The ULP model identifies significant locations by clustering. The example in Fig. 3 shows the use of a location clustering algorithm to identify significant locations. Next, a Markov model is built to predict the next location. Figure 6 shows a Markov model of the travel route between home, the Centennial Research Building (CRB), and the Department of Veterans Affairs (VA).

Occasionally Markov models are combined with recommendation methods to establish a location prediction system. For example, in the Collaborative Exploration and Periodically Returning (CEPR) model, Lian et al.<sup>[48]</sup> solved the problem of location prediction based only on context. Human movement patterns reveal the human preference for finding fresh locations, as do individual mobility trajectories. Lian et al.<sup>[48]</sup> proposed the CEPR model to predict user locations based on this human movement tendency. First, the authors studied unique characteristics in human movement behaviour and proposed a novel solution: the use of exploration prediction, for estimating whether



**Fig. 6 Examples of different location state transitions in a Markov model<sup>[51]</sup>.**

a visited location is new. Lian et al.<sup>[48]</sup> used a binary classification system to classify a movement behaviour as either an exploration or return to a visited location. Their experimental results had a 20% classification error rate on two check-in datasets, comprising  $6 \times 10^6$  (million) and  $3.6 \times 10^7$  records, respectively. Figure 7 shows the probability of return and the ratio of novel check-in to all check-in locations in the two datasets.

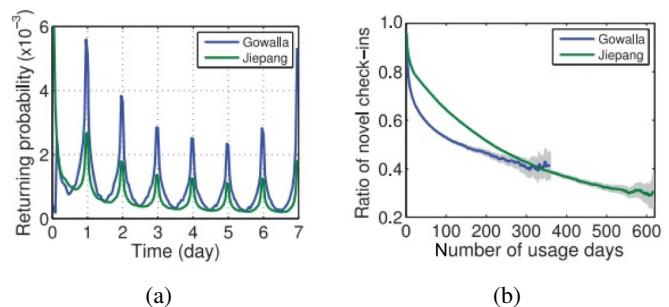
Figure 4 shows the correlation between novelty and various features. Using this exploration classification, the CEPR model integrates location prediction and location recommendation algorithms to generate a location predictor. A Markov model is used for location prediction and a location recommendation is used to solve the problem of predicting novel locations. Equation (3) below shows the CEPR framework, where  $Pr(Explore) \in [0, 1]$  is the exploration rate based on a binary classification,  $Pr(l)$  is the probability that a user arrives at a location, based on the prediction algorithm, and  $P_n(l)$  is the probability of a user exploring a novel location, based on the recommendation algorithm,  $Pr(l), P_n(l) \in [0, 1]$ .  $Pr(EXplore)$  appears to be a soft switch for controlling  $Pr(l)$  and  $P_n(l)$ .

$$P(l) = Pr(Explore)P_n(l) + (1 - Pr(Explore))Pr(l) \quad (3)$$

Content-based methods determine the correlation between locations. This method has two disadvantages: (1) if a user does not visit some locations, those locations do not appear in the user's prediction list and (2) similar content will predict the same location.

#### 4.1.2 Distribution-based methods

Distribution-based methods model user movement as a distribution with location and time as two random variables. In this method, the probability of a random variable is computed and then ranked to predict a

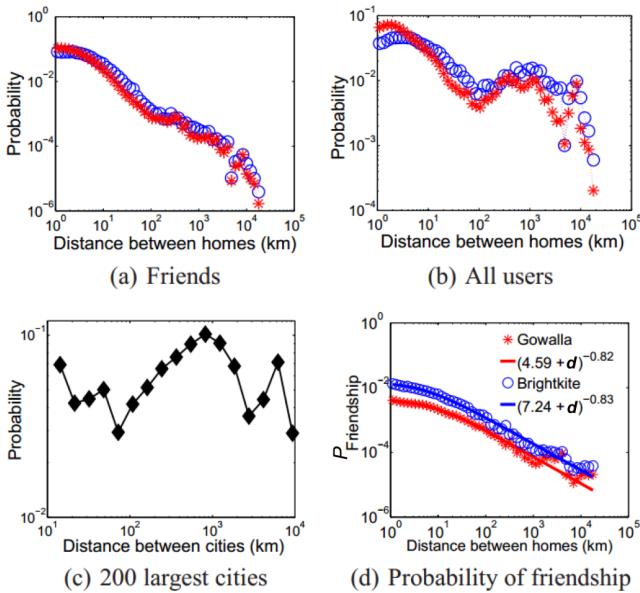


**Fig. 7 (a) Probability of return; (b) ratio of novel check-ins locations to all check-in locations in Gowalla and Jiepang datasets<sup>[48]</sup>.**

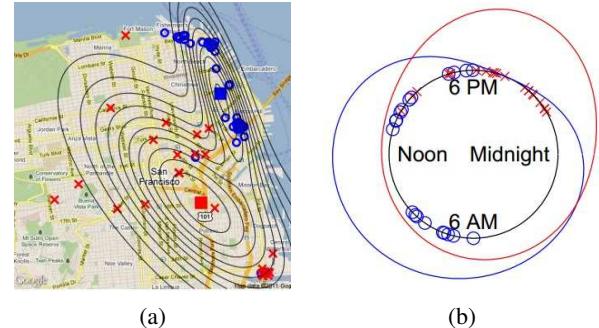
location.

Cho et al.<sup>[52]</sup> discovered that human movement in daily life is influenced by the factors of geographical limitations and social relations. In their collected data set, the authors identified a peculiar phenomenon whereby the social network structure has little effect on the spatial and temporal patterns of human short-range travel. Furthermore, social relationships account for just 10%–30% of people's daily life movement, and 50% of their periodic activity. The authors proposed a location prediction method based on these findings—the Periodic and Social Mobility Model (PSMM)—which describes human mobility based on periodic short-range travel and social network structures. The authors determined the home distance distributions of friends and all users, the distance between 200 large cities, and the probability of friendship as a function of distance, the results of which are shown in Fig. 8.

Figure 9 shows that the work place and home are the primary places in human daily life, and that most people visit their work place during the workweek daytime hours and their place of residence on work nights and weekend day times. Based on this finding, they proposed the Periodic Mobility Model (PMM) for predicting future user location states (based on partitions of all locations visited by users, both work- and home-related). First, PMM infers the geographic location centres of two latent locations for each user and



**Fig. 8 Distribution of distances between (a) homes of friends, (b) all users, (c) 200 large cities, and (d) the probability of friendship as a function of distance<sup>[52]</sup>.**



**Fig. 9 (a) Spatial model: geographic distribution of check-ins when at home or work place. (b) Temporal model: temporal distribution of check-ins when at home or work place, denoted by red and blue lines, respectively<sup>[52]</sup>.**

then models them based on their Gaussian distribution. Then, the PMM model determines the probability of whether a user is at home or the work place as a function of time of day.

Equations (4) and (5) are used to compute the probability of the user state distributions based on their Gaussian distributions:

$$\begin{aligned} N_H(t) &= \frac{P_{cH}}{\sqrt{2\pi\sigma_H^2}} \exp\left[-\left(\frac{\pi}{12}\right)\frac{(t-\tau_H)^2}{2\sigma_H^2}\right], \\ N_W(t) &= \frac{P_{cW}}{\sqrt{2\pi\sigma_W^2}} \exp\left[-\left(\frac{\pi}{12}\right)\frac{(t-\tau_W)^2}{2\sigma_W^2}\right] \end{aligned} \quad (4)$$

and then,

$$\begin{aligned} P[c_u(t) = H] &= \frac{N_H(t)}{N_H(t) + N_W(t)}, \\ P[c_u(t) = W] &= \frac{N_W(t)}{N_H(t) + N_W(t)} \end{aligned} \quad (5)$$

where  $\tau$  is the average time of day the user is at a location state, and  $\theta$  is the variance in time of day.

To integrate social information into the PMM, Cho et al.<sup>[52]</sup> introduced a check-in classification ( $z_u(t)$ ) to PMM.  $z_u(t) = 1$  to determine whether the check-in location is related to the user social network, and vice versa. The authors first determined the correlation between travel distance and social information, and then modeled the spatiotemporal pattern and social information to predict location. However, this method lacks any description of spatiotemporal patterns and social structure.

$$\begin{aligned} P[x(t) = x] &= P[x(t) = x | z_u(t) = 1]. \\ P[z_u(t) = 1] + P[x(t) = x | z_u(t) = 0] &= \\ P[z_u(t) = 0] \end{aligned} \quad (6)$$

Distribution-based methods model geographical characteristics and temporal information as a

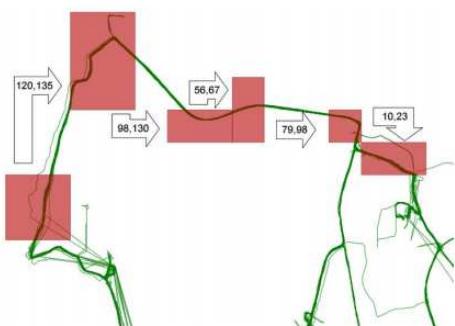
probability distribution model to compute the probability of users arriving at a location. The difficulty in distribution-based methods is that if a user's movement distributions do not fit the assumption, performance suffers.

#### 4.1.3 Pattern-based methods

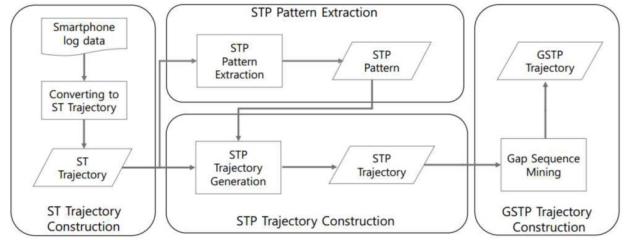
Trajectory pattern mining is another important branch of location-prediction research including methods based on sequential patterns<sup>[53,54]</sup>, frequency patterns<sup>[35]</sup>, integrated data patterns<sup>[55]</sup>, and periodic patterns<sup>[26]</sup>. These methods extract spatiotemporal patterns from trajectories to predict locations.

Here, we describe two predictor methods based on spatiotemporal patterns. Monreale et al.<sup>[56]</sup> proposed a dynamic pattern model called the T-pattern which is a dynamic mining method for extracting GPS trajectory data<sup>[53]</sup>. First, the authors used nearest neighbour methods to dynamically analyze the density distribution and extracted a range of interesting dense cells. Then, they constructed temporally annotated sequences and computed its relationship to the spatiotemporal pattern within a given time tolerance. Figure 10 shows an example of a T-pattern study. After extracting the T-pattern, Monreale et al.<sup>[56]</sup> stored it in a prefix-tree structure, called the T-tree. In the prediction process, this method computes a special score that includes the path and punctuality scores. Next, the space tolerance, time tolerance, path score, and punctuality score are integrated to make a prediction.

Lee et al.<sup>[57]</sup> also extracted spatiotemporal patterns to distinguish a T-pattern called the Gapped SpatioTemporal-Periodic (GSTP) pattern, whereby gapped sequence mining is used to build a pattern-based location prediction system. Figure 11 shows the training process of the proposed framework, for which the data source is smartphone log data. After data cleaning, the authors extracted the GSTP. This



**Fig. 10** Example of T-pattern in temporally annotated trajectory<sup>[53]</sup>.



**Fig. 11** Extracting process of spatiotemporal pattern and predictor training process<sup>[57]</sup>.

pattern extraction process involves computing the time range in which users stay at a unique place and the stay probability, based on the SpatioTemporal-Periodic (STP) pattern and use the GSTP to compute the time gap and transition probability between different locations. After constructing the GSTP trajectory, this trajectory is used to forecast the next location. This simple pattern describes the time gap and time period of a user's movement. The disadvantage of this pattern is that it is based only on observed spatial and temporal information and lacks any detailed description of movement patterns, such as the location visitation frequency.

In contrast to the GSTP, Monreale et al.<sup>[56]</sup> considered both time and space tolerance, and used a unique data structure to store the T-pattern, which is beneficial in predicting locations. Although pattern-based methods are common, the difficulty of extracting meaningful patterns is a non-trivial task. Pattern-based methods also seem overly exacting and lack diversity in visitation locations.

#### 4.1.4 Preference-based methods

Mobility preference is an important factor in predicting user location. Many studies have focused on user mobility preferences as a basis for predicting location. There are a number of methods available for determining user preference, with matrix-based methods being the most popular.

User location history can be used to generate a matrix and then matrix factorization can be used to capture user movement preferences. A tensor is an extension of the matrix. Bhargava et al.<sup>[58]</sup> used tensor factorization methods in multi-dimensional collaborative filtering to predict location via user profiles, user's short message in social network, and user location and temporal information. Human movement is related to user preferences, user activities, and user spatiotemporal patterns. The use of additional information enables the

improvement of the overall prediction outcome. Multi-source data fusion and collaborative filtering are key techniques. To solve the sparsity problem in check-in data, Bhargava et al.<sup>[58]</sup> proposed tensor factorization to complete multi-dimensional collaborative filtering tasks. They jointly analyzed the constructed tensor and matrices and formulated an objective function as shown in Eq. (7).

$$\begin{aligned} \text{Function} = & \frac{1}{2} \|W \times (X - U \circ L \circ A \circ T)\|^2 + \\ & \frac{\lambda_1}{2} \|Y - LA^T\|^2 + \frac{\lambda_2}{2} \|S - AA^T\|^2 + \\ & \frac{\lambda_3}{2} \|Z - LL^T\|^2 + \frac{\lambda_4}{2} \|TR\|^2 + \\ & \frac{\lambda_5}{2} (\|U\|^2 + \|L\|^2 + \|A\|^2 + \|T\|^2) \end{aligned} \quad (7)$$

where  $W$  is a weight tensor,  $X$  is a tensor,  $U$  is a user matrix,  $L$  is a location matrix,  $A$  is an activity matrix, and  $T$  is a time matrix.  $\|\cdot\|^2$  is Frobenius norm,  $\circ$  is the outer product, and  $\lambda_1 - \lambda_5$  are the model parameters. This method integrates user, activity, location, and temporal information to predict location by tensor factorization. The drawback of this method is that it is typically both resource and time consuming.

Single type mobility datasets often contain limited information. For example, check-in datasets include only check-in behaviour and social relationship. Heterogeneous mobility datasets include bus, taxi, daily life mobility, and so on. The use of various data sources provides new perspectives regarding movement patterns. Wang et al.<sup>[47]</sup> proposed a Regularity Conformity Heterogeneous (RCH) model for analyzing heterogeneous datasets and used a gravity model to determine the spatial influence. The RCH model assumes human movement to be affected by both regularity term and conformity term, and splits the geospatial space into many grid cells, as shown in Fig. 12.

Figure 12 associates each venue with a grid cell, in which a plus sign means that the user has visited this venue. RCH computes the probability of user  $u_i$  visiting  $v_j$ , denoted as  $Pr(v_j|u_i)$ , based on three factors. The first factor is the visiting frequency of a grid cell as shown in Figs. 12a and 12c. The second factor is the transition probability of the geospatial influence between different grid cells as determined by the gravity model. Figures 13a and 13b show the spatial influences of a bar district and an IT

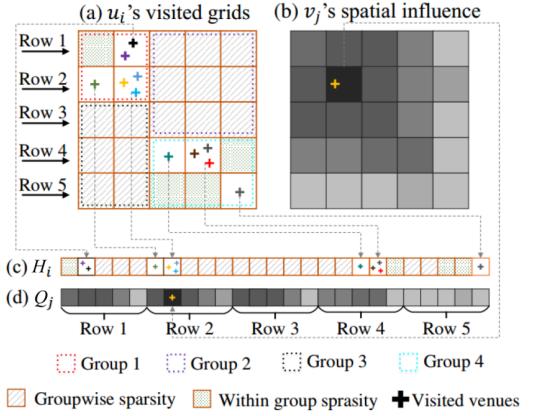


Fig. 12 Method for determining regularity term<sup>[47]</sup>.

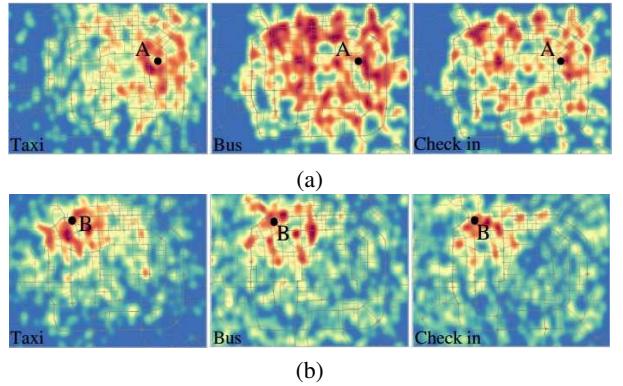


Fig. 13 Spatial influence of a location using a gravity model based on taxi, bus, and check-in datasets. (a) Point A is a bus district, (b) point B is an IT district<sup>[47]</sup>.

district, respectively. In contrast to traditional methods, the gravity model enriches the life and commuting information via training by heterogeneous trajectory data, such as taxi, bus, and check-in data. The third factor is the venue visiting frequency. The regularity term is computed using Eq. (8) below.

$$\begin{aligned} Pr(v_j|u_i) \propto & \sum_{k=1}^I Pr(d_k|u_i) Pr(v_j|d_k) = \\ & \sum_{k=1}^I Pr(d_k|u_i) Pr(d_{kj}|d_k) Pr(v_j|d_{kj}) \end{aligned} \quad (8)$$

RCH describes conformity term using a time changing matrix factorization model, as shown in Eq. (9):

$$R_{i,j}^{(c)}(t) = (U_i + U_i(t)) V_j^T \quad (9)$$

By incorporating regularity term and conformity terms, the RCH model is defined as follows.

$$R_{i,j}(t) = R_{i,j}^r(t) + R_{i,j}^c(t) \quad (10)$$

The final objective function of RCH is that expressed by Eq. (11). By modeling the regularity term and conformity term, RCH considers not only individual preferences based on regularity term, but also the influence of others who are similar to the user. Then, it incorporates heterogeneous mobility data for the first time into its spatial influence analysis.

$$\begin{aligned} \min_{U, U(t), V} & \sum_{t \in \mathcal{T}} \|R(t) - (U + U(t)V^T)\|_F^2 + \\ & \gamma(\|U\|_F^2 + \|V\|_F^2) + \beta \sum_{t \in \mathcal{T}} \|U(t)\|_F^2 \end{aligned} \quad (11)$$

The above approaches capture user or public preferences via matrix or tensor factorization. Like pattern-based methods, preference learning can extract user movement behaviour. However, their disadvantages are also similar. Most preference-based methods ignore the evolution of preference and the randomness of mobility.

#### 4.1.5 Social relation-based methods

Social relation-based methods utilize social relations to model user movement, and then use these relations to infer future visiting locations of a given user. Previous works have often depended on historically visited locations. However, when user data contains few previous locations, prediction methods cannot generate any useful result. The social relation feature can be used to alleviate the effect of cold starts and the data parsing problem in location prediction models.

Gao et al.<sup>[59]</sup> proposed a geo-social correlation (gSCorr) model to solve the cold start problem based on social information. The GSCorr model determines the correlation between social network and geographical distance, and defines a complex matrix of four relationships between social information and distance, as listed in Table 3, where  $F$  is a friend in a social network,  $\bar{F}$  is not an observed friend in a social network,  $D$  is a long geographical distance, and  $\bar{D}$  is a short distance.  $S_{FD}$  describes a user's friends in a social network who live a short distance away. Figure 14 shows a user's check in behaviour in different social correlation aspects. GSCorr predicts the probability of a user checking-in at a new location  $P_u^t(l)$  (Eq. (12)), where  $\phi_1 - \phi_4$  are four distributions that govern

Table 3 Geo-social correlation<sup>[59]</sup>.

	$F$	$\bar{F}$
$\bar{D}$	$S_{F\bar{D}}$ : LocalFriends	$S_{\bar{F}\bar{D}}$ : LocalNon-friends
$D$	$S_{FD}$ : DistantFriends	$S_{\bar{F}D}$ : DistantNon-friends

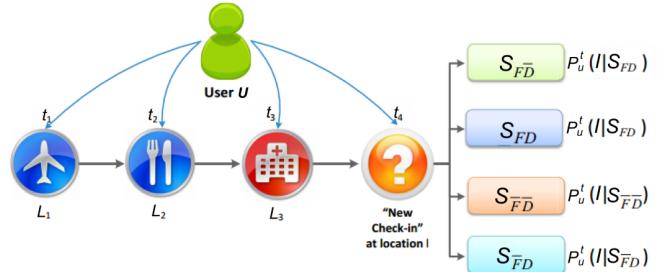


Fig. 14 Influence of new check-in behaviour on geo-social correlation<sup>[59]</sup>.

the strength of diversity in the geo-social correlation,  $P_u^t(l|S_x)$  is the probability of a user checking-in at a location in the gSCorr model, and  $S_x$  is a user's geo-social circle.

$$P_u^t(l) = \phi_1 P_u^t(l|S_{F\bar{D}}) + \phi_2 P_u^t(l|S_{\bar{F}\bar{D}}) + \phi_3 P_u^t(l|S_{FD}) + \phi_4 P_u^t(l|S_{\bar{F}D}) \quad (12)$$

In addition to studying the correlation between social circle and geographical distance, Guo et al.<sup>[60]</sup> also paid attention to determined check-in patterns based on check-in history. The authors obtained user check-in patterns based on the power-law distribution and short-term effect to explain user check-in behaviour in a social network. To do so, they utilized the influence of social and historical ties on user check-ins. They assumed (1) user check-in behavior to follow a power-law distribution and that (2) the check-in history trajectory has a short-term effect. Based on these assumptions, they proposed a social historical model that uses a Hierarchical Pitman Yor (HPY) process<sup>[61]</sup> to model a user's historical check-in sequence of LBSNs. The HPY extracts the power-law distribution and short term effect and formulates the user's next check-in location based on that user's check-in history, as shown in Eq. (13), where  $G$  is the next check-in location,  $d \in [0, 1]$  is a discount parameter for controlling the power-law property,  $\gamma$  is a parameter, and  $G_0$  is the base distribution<sup>[61]</sup>. The parameters in Eq. (13) can be inferred based on the observed check-in history.

$$G \sim PY(d, \gamma G_0) \quad (13)$$

To introduce social ties to the model, parameter  $\eta$  is introduced to denote social influence. Equation (14) is a social historical model that integrates social ties, where  $P_H^i(c_{n+1} = l)$  is the probability of user check-in at location  $l$  based on the history trajectory,  $P_S^i(c_{n+1} = l)$  is the probability of check-in at location  $l$  based on user social ties, and  $i$  is the user  $u_i$ :

$$P_{SH}^i(c_{n+1} = l) = \eta P_H^i(c_{n+1} = l) +$$

$$(1 - \eta) P_S^i (n+1 = l) \quad (14)$$

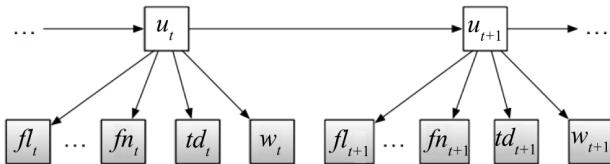
where we can compute the probability of social ties  $P_S^i (n+1 = l)$  using Eq. (15).  $\mathcal{N}(u_i)$  is the friend set of a user  $u_i$ , and  $P_{HPY}^{i,j}(c_{n+1} = l)$  is the probability of a user  $u_i$  check-in location  $l$ , where the probability is computed using the HPY process and user check-in history. Equation (16) is used to compute  $P_H^i(c_{n+1} = l)$ .

$$P_S^i (c_{n+1} = l) = \sum_{u_j \in \mathcal{N}(u_i)} sim(u_i, u_j) P_{HPY}^{i,j}(c_{n+1} = l) \quad (15)$$

$$P_H^i(c_{n+1} = l) = P_{HPY}^{i,j}(c_{n+1} = l) \quad (16)$$

Sadilek et al.<sup>[62]</sup> proposed the Friendship and location analysis and prediction (Flap) model to infer user location despite users not having released their private data. The authors built a Bayesian network for modeling the effect of a user's friends movement patterns to realize two functions. First, Flap can predict the social relationships between user movement patterns. Since Bayesian networks are a probabilistic, Flap is a graph network based on probability inference. Second, the use of social relationship improves the results of user location prediction. Sometimes, despite the fact that some users keep their locations private, Flap can infer their location based on the location information of their friends, which is helpful in solving "cold start" problems. Flap predicts location using a sequence of visited locations and temporal information about a user's friends. Flap output is also a sequence of locations that a user has visited over a given period of time. Figure 15 shows a dynamic Bayesian network for modeling locations visited by a user. Flap learns and infers the parameters in a dynamic Bayesian network and then predicts user locations.

Social relation-based methods focus on the effect of social relations on location prediction. These methods can also alleviate the problems of cold start and data sparsity. Social relations represent additional information that can help improve location prediction



**Fig. 15** Two consecutive time slices of a Flap dynamic Bayesian network for modeling user movement pattern based on  $n$  friends<sup>[62]</sup>.

performance.

#### 4.1.6 Time-dependent methods

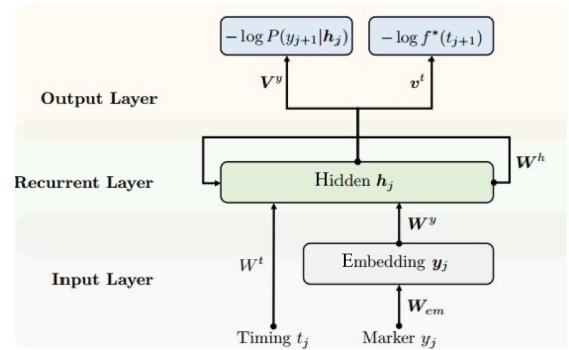
Most location prediction algorithms focus on geography or social characteristics, while ignoring temporal information. Some algorithms build a time-dependent model, but using a stochastic process is a better choice. Stochastic process models utilize time factors as random variables and embed location factors into the stochastic process. A typical example is the point process.

Du et al.<sup>[63]</sup> proposed the Recurrent Marked Temporal Point Process (RMTPP) to simultaneously model visiting time and location. The basic concept of RMTPP is to model movement history using a nonlinear function. RMTPP also uses a recurrent neural network to automatically learn a representation of influences from a user mobility history. Figure 16 shows the RMTPP framework.

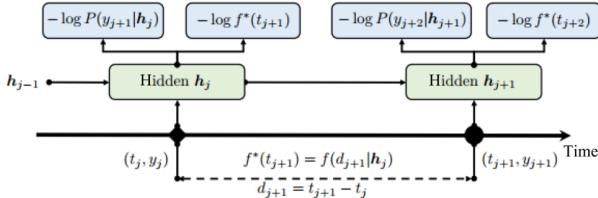
For a given check-in trajectory  $T = (t_j, l_j)_{j=1}^n$  at the  $j$ -th check-in, location  $l_j$  is first embedded into a latent space. Next, embedded vector and temporal information are fed into the recurrent layer and RMTPP automatically learns a representation of the visiting history. The output layer infers the next location and time depending on the representation. The key RMTPP process occurs in the recurrent layer, which undergoes a recurrent temporal point process. Figure 17 shows the recurrent temporal point process, which obtains a general representation of the fuzzy relation between temporal and spatial information.

Equation (17) shows the hidden layer, which has three parts. Here,  $y_j$  is the location information,  $t_j$  is the temporal information,  $W^*$  is the embedded matrix,  $h_j$  is the representation of the  $j$ -th check-in event, and  $b_h$  is the base term.

$$h_j = \max\{W^y y_j + W^t t_j + W^h h_{j-1} + b_h, 0\} \quad (17)$$



**Fig. 16** RMTPP framework<sup>[63]</sup>.



**Fig. 17** Two consecutive time slices in the hidden layer of the recurrent marked temporal point process<sup>[63]</sup>.

The core of this model is the hidden layer representation and conditional intensity function of the temporal point process  $\lambda^*(t)$ , as shown in Eq. (18). The conditional intensity function consists of three parts, past influence, current influence, and base intensity. RMTTP uses  $\lambda^*(t)$  to model user check-in behaviour with temporal information being the key factor, thereby differing from spatial-based algorithms.

$$\lambda^*(t) = \exp\left(\underbrace{v^T h_j}_{\text{Past influence}} + \underbrace{w^t(t - t_j)}_{\text{Current influence}} + \underbrace{b^t}_{\text{Base intensity}}\right) \quad (18)$$

In addition, Zarezade et al.<sup>[64]</sup> proposed a probabilistic model based on point process, in which a periodic kernel function is used to capture the user time period and multi-nominal distribution of locations. The whole framework is similar that of the RMTPP. Point processes are effective in modeling time-varying processes. To date, blending location prediction into the temporal point process is a newly emerging research direction.

In contrast with traditional models, time-dependent methods model time as an important factor in location prediction. Although prediction performance can be improved in this way, the correlation between spatial and temporal information is barely considered.

#### 4.1.7 Representation-based methods

Traditional trajectory representation uses a point sequence, and some researchers have proposed new trajectory representation methods such as the extraction of trajectory features and representation based on deep learning.

Noulas et al.<sup>[65]</sup> regarded the location prediction problem as a ranking problem, whereby every check-in associated location and time is defined as a tuple  $l, t$ , ranking at the highest possible location in a user historical list of visited locations. Prediction features can be classified as user mobility, global mobility, or temporal features.

- The user mobility feature includes historical visits

and categorical preferences.

- The global mobility feature includes the popularities of geographic and rank distances.
- The temporal feature captures both the time of the location visit and the temporal patterns associated with significant locations.

Using these features, Noulas et al.<sup>[65]</sup> proposed a rank model for predicting future locations. Location prediction ranking methods comprise supervised models that obtain the check-in data tuple position in the history list. This model extracts features from check-in datasets, thereby providing a new way to consider the location prediction problem. However, it ignores the spatiotemporal sequence and sparsity characteristic of check-in data.

In recent years, deep learning techniques have been developed to a great extent, enjoy wide application<sup>[66]</sup>, and typically yield good performance. To date, many researchers predict location using deep learning techniques<sup>[49, 67, 68]</sup>. For example, Liu et al.<sup>[49]</sup> have extended Recurrent Neural Networks (RNNs) to model time and location. Specifically, they used spatiotemporal RNNs (ST-RNNs) to predict location.

$$h_{t,q_t}^u = f\left(\sum_{q_{t_i}^u \in Q^u, 0 < t_i < t} S_{q_t^u - q_{t_i}^u} T_{t-t_i} q_{t_i}^u + Ch_0^u\right) \quad (19)$$

Equation (19) is a representation of user  $u$  at time  $t$ , where  $S$  is a distance-specific transition matrix,  $q$  is a latent vector of the location of user visits,  $C$  is a recurrent connection, and  $h_0$  is the initial status. The hidden layers in the ST-RNN integrate the effect of the geographic distances between locations and visiting history. ST-RNN models consider geographic distance in RNNs to predict location. Deep learning techniques are widely used in location prediction. Zhang et al.<sup>[67]</sup> proposed a deep learning method for predicting citywide crowd flows. First, the authors divided the city into grid cells, then they constructed an end-to-end deep spatiotemporal residual network, called the ST-ResNet, to predict the inflow and outflow of grid cells in a city. ST-ResNet consists of four major components that model the external, trend, period, and closeness influences of a grid cell. ST-ResNet determines their effects based on the inflow and outflow of grid cells in a city and then trains the network to predict crowd flows in a given location.

Representation-based methods explore new trajectory representations to predict location. However, finding a good representation to capture the

characteristics of trajectory data remains a challenge.

#### 4.1.8 Semantic-based methods

The aforementioned approaches often focus on spatiotemporal space, however, few researchers have paid attention to the trajectory of semantic space. Semantic-based predictors enable better reasoning and therefore better location prediction results. Ying et al.<sup>[23]</sup> proposed a semantic framework for location prediction via semantic pattern mining, called SemanPredict. First, the authors extracted frequent locations in a user's movement history and the semantic information associated with these movements, and then they generated semantic trajectory patterns. SemanPredict then generates two tree structures to store these patterns. In the prediction process, the score of the tree structure path is computed to predict location. Semantic-based methods enable a better understanding of the semantic information associated with visits to locations.

### 4.2 Model categorization

Model categorization refers to the focus of an approach being a single movement object or a group. This categorization provides another perspective for distinguishing different location prediction algorithms.

#### 4.2.1 Single-object models

The earliest location prediction studies investigated single object movement, like the Markov chain (Song et al.<sup>[43]</sup>, Ashbrook and Starner<sup>[51]</sup>), spatiotemporal pattern mining (Bhargava et al.<sup>[58]</sup>, Lee et al.<sup>[19]</sup>, Monreale et al.<sup>[56]</sup>), and ranking problems (Noulas et al.<sup>[69]</sup>). These methods determine the location transition probability, user movement preference, and movement patterns with respect to single-object movement<sup>[70]</sup>. Although user future locations are mainly dependent on the user himself, other factors also play a vital role in location prediction, such as social information, group movement tendencies, and geographic distance, to mention a few, and these factors are not considered in single object spatiotemporal data. Also, single-object models lack robustness and are sensitive to anomalous and sparse data. Since a single object can be characterized by random movement and usually produces sparse data, location transition probability and movement patterns cannot effectively model real individual movement patterns.

#### 4.2.2 Group models

Unlike single-object models, group models primarily

study a group of moving objects based on the hypothesis that both human beings and animals have social relationships. Group model researchers consider movement behavior to "follow the crowd" to some degree. Typical group methods are based on clustering (Ashbrook and Starner<sup>[51]</sup>), matrix analysis (Wang et al.<sup>[47]</sup>), and social relation analysis (Cho et al.<sup>[52]</sup>, Sadilek et al.<sup>[62]</sup>). The goal of these methods is to extract group movement tendencies or patterns to enable location prediction. Cluster-based methods can be divided into two types: (1) determining user clusters to explore user social relations, and (2) identifying significant locations based on the data of all users. Both types determine the correlation between users and others. Matrix analysis techniques employ matrix factorization and collaborative filtering to obtain common patterns or relationships between users. Social relation information is direct data that reveals a user's social structure. The researcher builds a social graph to analyze the influence of social information on user movement. Group models often study external influences on users, such as social information, geographic distance, and even the effect of events. These additional types of information can enable improved location prediction performance. Unlike single-object models, group models can reveal the movement of a group of users in some scenarios, and have attracted interest by governmental organizations for optimizing traffic flow and city planning.

#### 4.2.3 Hybrid models

Both single-object and group models have advantages and disadvantages. Some researchers have proposed hybrid models that integrate the two model types. For example, Wang et al.<sup>[47]</sup> modeled regularity and conformity terms, respectively, and added these two factors to predict location. Lian et al.<sup>[48]</sup> modeled the correlation between periodic user behaviour and novel exploration behaviour. Sadilek et al.<sup>[62]</sup> learned user's social information in social networks to enhance the performance of location prediction. Their model integrates a user's profile, activity (based on an analysis of a short message), temporal information, and location to build a predictor model. Generally, hybrid mainly modes have three main concerns: (1) developing single-object and group models, (2) correlating or integrating the single-object and group models, and (3) using additional information to improve the performance of their prediction systems. Recently, hybrid models have become the keystone in location-predication research due to their excellent performance.

### 4.3 Summary

Table 4 provides an overview of some representative approaches with respect to the type of method used, model category, and the data source type.

## 5 Location Prediction Evaluations

In this section, we introduce some common trajectory datasets, including GPS, check-in, and Wi-Fi datasets. Next, we present some evaluation matrices for assessing location prediction.

### 5.1 Datasets

In recent years, there has been an increasing number of public trajectory datasets, including GPS, bike sharing, check-in, and Wi-Fi data. Zheng<sup>[11]</sup> and Bao et al.<sup>[73]</sup> have provided summaries of these data. We summarize and supplement these datasets in Table 5.

*BrightKite check-in dataset* is provided by a location based social network service provider. This dataset includes 4 491 143 check-ins from 58 228 users and their social relation data (4 491 143 relations) over the period from Apr. 2008 – Oct. 2010<sup>[52]</sup>.

*Gowalla check-in dataset* is provided by a location based social network website in which users share their locations by checking-in<sup>[52]</sup>. This dataset comprises a total of 6 442 890 check-ins from 196 591 users over the

period from Feb. 2009 – Oct. 2010.

*Foursquare check-in datasets* include three data sets: (1) Foursquare 1 is simply a check-in dataset with 12 000 000 check-ins from 679 000 users<sup>[74]</sup>; (2) Foursquare 2 includes 221 128 check-ins from 49 062 users in New York City and 104 478 check-ins from 31 544 users in Los Angeles. As such, this dataset consists of both check-in information as well as social relation, and user and venue profiles<sup>[75]</sup>; (3) Foursquare 3 is similar to the Foursquare 2 dataset with respect to the data types, and contains a total of 33 596 users<sup>[60]</sup>.

*GeoLife trajectory dataset*, collected by Microsoft Research Asia, includes 17 621 trajectories from 183 users' life trajectory datasets over a five-year period (from April 2007 to August 2012)<sup>[21, 76]</sup>. This dataset is useful in analyses of user long-term movement patterns.

*Porto taxi trajectory dataset* contains a complete year (from 01/07/2013 to 30/06/2014) of the 1 710 670 trajectories of 442 taxis, the details of which can be obtained from the website (<https://www.kaggle.com/c/pkdd-15-predict-taxi-service-trajectory-i>).

*Bike sharing dataset* contains 3 million trajectory records generated by 300 000 users and 400 000 bikes. For more details, readers are referred to the website (<https://biendata.com/competition/mobike/>).

*Animal movement dataset* contains the radio-

**Table 4 Summary of existing location prediction algorithms. (Social relation based prediction algorithm: S-R; time dependent based prediction algorithm: T-D; active recording trajectory data: A; and passive recording trajectory data: P)**

	Method				Model categorization			Data		
	Content	Distribution	Pattern	Preference	S-R	T-D	Representation	Singleness	Group	Hybrid
Song et al. <sup>[43]</sup>	✓							✓		P
Ashbrook and Starner <sup>[51]</sup>	✓								✓	A
Lian et al. <sup>[48]</sup>	✓							✓		A
Xue et al. <sup>[71]</sup>	✓							✓		P
Cho et al. <sup>[52]</sup>		✓							✓	A
Lee et al. <sup>[57]</sup>			✓					✓		P
Monreale et al. <sup>[56]</sup>			✓					✓		P
Bhargava et al. <sup>[58]</sup>				✓					✓	A
Wang et al. <sup>[47]</sup>				✓					✓	A
Gao et al. <sup>[59]</sup>					✓				✓	A
Gao <sup>[60]</sup>					✓				✓	A
Du et al. <sup>[63]</sup>						✓		✓		A
Zarezade et al. <sup>[64]</sup>						✓		✓		A
Noulas et al. <sup>[65]</sup>							✓	✓		A
Liu et al. <sup>[49]</sup>							✓	✓		A
Zhang et al. <sup>[67]</sup>						✓			✓	P
Zhang et al. <sup>[68]</sup>						✓		✓		P
Ying et al. <sup>[23]</sup>							✓	✓		P
Ying et al. <sup>[72]</sup>							✓	✓		P

**Table 5** Summary of trajectory datasets. A: active recording date, P: passive recording data.

Dataset	Source	Type
BrightKite dataset	Check-in & Social information	A
Gowalla dataset	Check-in & Social information	A
Foursquare datasets	Check-in & Social information	A
GeoLife dataset	GPS trajectory of human life	P
Proto taxi dataset	GPS trajectory of taxi	P
Bike sharing dataset	GPS trajectory of bike sharing	P
Animal movement dataset	GPS trajectory of animal	P
Hurricane tracking dataset	GPS trajectory of Hurricane	P

telemetry locations (along with other information) of elk, deer, and cattle from 1993 through 1996. For more the details, readers are referred to the website (<https://www.movebank.org/>). This dataset is useful in studies of animal behaviour in the wild.

*Hurricane tracking dataset* contains 1740 trajectories of Atlantic hurricanes over the 1851 to 2012 period, as provided by the U.S. National Hurricane Service (NHS)<sup>[19]</sup>. This dataset is used in meteorological and natural disaster prevention studies.

## 5.2 Evaluation matrices

Location prediction evaluation mainly focuses on precision, rank, and recall. Precision is based on the accuracy@k (top  $k$  locations) and Mean Average Precision (MAP), rank is determined based on the Normalized Discounted Cumulative Gain (NDCG) and Average Precision Rank (APR), recall is typically described by an F1 score.

*Accuracy@k* is an index for evaluating the top- $k$  prediction locations whether or not they represent the real locations.

*MAP* is determined based on an information retrieval field. This value is the MAP, as defined in Eq. (20), where  $N$  is the number of locations and  $rel(r)$  is a binary function related to the relevance of a given location rank.

$$MAP = \frac{\left( \sum_{r=1}^n (P(r) \cdot rel(r)) \right)}{N} \quad (20)$$

*APR* is the average percentile rank of a prediction for location  $l_j$ <sup>[47]</sup>. *PR* is defined as shown in Eq. (21), where  $rank(l_j)$  is the position of location  $l_j$  and  $N$  is the number of locations.

$$PR = \frac{N - rank(l_j) + 1}{N} \quad (21)$$

*NDCG* is an index of information retrieval, which is used to evaluate the quality of a ranking. For more

details, readers are referred to the book *Introduction to Information Retrieval*<sup>[77]</sup>.

*F1 score* is an index that considers precision and recall, as shown in Eq. (22), where

$$\begin{aligned} precision &= \frac{\text{number of real prediction locations}}{\text{total number of prediction locations}}, \\ recall &= \frac{\text{number of real prediction locations}}{\text{total number of ground truth locations}}, \\ F_1 &= 2 \cdot \frac{precision \cdot recall}{precision + recall} \end{aligned} \quad (22)$$

## 6 Applications and Future Work

In this section, we discuss the applications of location prediction and future work.

### 6.1 Applications

Location prediction has wide application and location predictors play an important role in urban management. These models can help alleviate traffic congestion, improves urban governmental planning<sup>[78–81]</sup>, and help predict crime rates<sup>[82]</sup>. Location prediction can also help advertisers to promote sales based on location. LBSN based data mining, which builds a real relation between on-line social networks and real life<sup>[83–85]</sup>, is used to recommend friends or locations and detect anomalous check-in locations. Location prediction is currently in a period of rapid development, and various techniques are being used based on the applications at hand.

### 6.2 Future work

Location prediction has been the subject of study for many years. However, it continues to face many challenges and problems. Future location prediction systems may have two thrusts: (1) correlating spatial and temporal information, and (2) developing effective and efficient prediction systems.

**Correlating spatial and temporal information** is a key problem in trajectory data mining. Current methods typically consider the spatial and temporal spaces independently. New solutions must be found to solve this problem. Examples currently include the use of point processes and RNNs. Another approach for solving this problem is to identify new representations.

**Effectiveness and efficiency** are two key and ongoing factors in location prediction systems. The earliest methods were built using small datasets for the purposes of validating the methods. The complexity of the methods and limitations in computer resources constrain the use of prediction systems involving large-

scale matrix factorization and sparse data in LBSNs. Effective and efficient location prediction systems are critical, as is the application of predictors to large volumes of spatiotemporal data.

## 7 Conclusion

In this article, we provided an overview of location prediction ranging from trajectory data preprocessing to forecasting location and the evaluation of location prediction systems. First, we introduced the basic concepts of location prediction, the different types of data sources, the challenges associated with location predictions and the location prediction framework. We introduced trajectory data preprocessing methods and then identified the classification of location prediction model types and discussed these models in detail. Next, we categorized location-prediction models as either single-object or group models or shared insights about these approaches. We also listed the available public datasets and evaluation methods to help readers conduct their own research. Lastly, we discussed location-prediction applications and future work.

## Acknowledgment

We thank the editors and reviewer for everything you have done for us. This research was supported by the Foundation of Science & Technology Department of Sichuan Province (Nos. 2017JY0027 and 2016GZ0075), the National Key Research and Development Program (2016YFB0502300).

## References

- [1] Y. L. Wang, Y. Zheng, and Y. X. Xue, Travel time estimation of a path using sparse trajectories, in *Proc. 20<sup>th</sup> ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, New York, NY, USA, 2014, pp. 25–34.
- [2] H. K. Pao, J. Fadlil, H. Y. Lin, and K. T. Chen, Trajectory analysis for user verification and recognition, *Knowl.-Based Syst.*, vol. 34, pp. 81–90, 2012.
- [3] J. Yuan, Y. Zheng, X. Xie, and G. Z. Sun, Driving with knowledge from the physical world, in *Proc. 17<sup>th</sup> ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, San Diego, CA, USA, 2011, pp. 316–324.
- [4] A. L. Barabási, The origin of bursts and heavy tails in human dynamics, *Nature*, vol. 435, no. 7039, pp. 207–211, 2005.
- [5] D. Brockmann, L. Hufnagel, and T. Geisel, The scaling laws of human travel, *Nature*, vol. 439, no. 7075, pp. 462–465, 2006.
- [6] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, F. F. Li, and S. Savarese, Social LSTM: Human trajectory prediction in crowded spaces, in *IEEE Conf. Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 2016, pp. 961–971.
- [7] F. Wu and Z. H. Li, Where did you go: Personalized annotation of mobility records, in *Proc. 25<sup>th</sup> ACM Int. Conf. Information and Knowledge Management*, Indianapolis, IN, USA, 2016, pp. 589–598.
- [8] C. L. Wang, D. De, and W. Z. Song, Trajectory mining from anonymous binary motion sensors in Smart Environment, *Knowl.-Based Syst.*, vol. 37, pp. 346–356, 2013.
- [9] W. Y. Zhu, W. C. Peng, L. J. Chen, K. Zheng, and X. F. Zhou, Modeling user mobility for location promotion in location-based social networks, in *Proc. 21<sup>th</sup> ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, Sydney, Australia, 2015, pp. 1573–1582.
- [10] M. Q. Lv, L. Chen, Y. B. Shen, and G. C. Chen, Measuring cell-id trajectory similarity for mobile phone route classification, *Knowl.-Based Syst.*, vol. 89, pp. 181–191, 2015.
- [11] Y. Zheng, Trajectory data mining: An overview, *ACM Trans. Intell. Syst. Technol.*, vol. 6, no. 3, p. 29, 2015.
- [12] J. J. Liu, K. Zhao, P. Sommer, S. Shang, B. Kusy, J. G. Lee, and R. Jurdak, A novel framework for online amnesic trajectory compression in resource-constrained environments, *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 11, pp. 2827–2841, 2016.
- [13] K. F. Richter, F. Schmid, and P. Laube, Semantic trajectory compression: Representing urban movement in a nutshell, *J. Spat. Inf. Sci.*, vol. 2012, no. 4, pp. 3–30, 2012.
- [14] Y. T. Jia, Y. Z. Wang, X. L. Jin, and X. Q. Cheng, Location prediction: A temporal-spatial Bayesian model, *ACM Trans. Intell. Syst. Technol.*, vol. 7, no. 3, p. 31, 2016.
- [15] W. C. Lee and J. Krumm, *Trajectory Preprocessing*. New York, NY, USA: Springer, 2011.
- [16] J. Muckell, J. H. Hwang, V. Patil, C. T. Lawson, F. Ping, and S. S. Ravi, SQUISH: An online approach for GPS trajectory compression, in *Proc. 2<sup>nd</sup> Int. Conf. Computing for Geospatial Research & Applications*, Washington, DC, USA, 2011, p. 13.
- [17] J. Muckell, P. W. Olsen, J. H. Hwang, C. T. Lawson, and S. S. Ravi, Compression of trajectory data: A comprehensive evaluation and new approach, *Geoinformatica*, vol. 18, no. 3, pp. 435–460, 2014.
- [18] Q. Gao, F. L. Zhang, R. J. Wang, and F. Zhou, Trajectory big data: A review of key technologies in data processing, *J. Softw.*, vol. 28, no. 4, pp. 959–992, 2017.
- [19] J. G. Lee, J. W. Han, and K. Y. Whang, Trajectory clustering: A partition-and-group framework, in *Proc. 2007 ACM SIGMOD Int. Conf. Management of Data*, Beijing, China, 2007, pp. 593–604.
- [20] S. Tiwari and S. Kaushik, Popularity estimation of interesting locations from visitor's trajectories using fuzzy inference system, *Open Comput. Sci.*, vol. 6, no. 1, pp. 8–24, 2016.

- [21] Y. Zheng, L. Z. Zhang, X. Xie, and W. Y. Ma, Mining interesting locations and travel sequences from GPS trajectories, in *Proc. 18<sup>th</sup> Int. Conf. World Wide Web*, Madrid, Spain, 2009, pp. 791–800.
- [22] C. Brunsdon, *Computing with spatial trajectories*, *Int. J. Geogr. Inf. Sci.*, vol. 27, no. 1, pp. 208–209, 2013.
- [23] J. J. C. Ying, W. C. Lee, T. C. Weng, and V. S. Tseng, Semantic trajectory mining for location prediction, in *Proc. 19<sup>th</sup> ACM SIGSPATIAL Int. Conf. Advances in Geographic Information Systems*, Chicago, IL, USA, 2011, pp. 34–43.
- [24] C. S. Jensen and N. Tradišauskas, Map matching, in *Encyclopedia of Database Systems*, L. Liu and M. T. Özsu, eds. Springer, 2009, pp. 1692–1696.
- [25] J. S. Greenfeld, Matching GPS observations to locations on a digital map, in *Proc. 81<sup>st</sup> Annual Meeting of the Transportation Research Board*, Washington, DC, USA, 2002.
- [26] H. B. Yin and O. Wolfson, A weight-based map matching method in moving objects databases, in *Proc. 16<sup>th</sup> Int. Conf. Scientific and Statistical Database Management*, Santorini Island, Greece, 2004, pp. 437–438.
- [27] H. Y. Li, Y. Ge, R. C. Hong, and H. S. Zhu, Point-of-interest recommendations: Learning potential check-ins from friends, in *Proc. 22<sup>nd</sup> ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, San Francisco, CA, USA, 2016, pp. 975–984.
- [28] Y. Zheng, Location-based social networks: Users. in *Computing with Spatial Trajectories*, Y. Zheng and X. F. Zhou, eds. Springer, 2011, pp. 243–276.
- [29] C. Barrios, H. Himberg, Y. Motai, and A. Sad, Multiple model framework of adaptive extended Kalman filtering for predicting vehicle location, in *2006 IEEE Intelligent Transportation Systems Conference*, Toronto, Canada, 2006.
- [30] J. Hightower and G. Borriello, Particle filters for location estimation in ubiquitous computing: A case study, in *Proc. 6<sup>th</sup> Int. Conf. Ubiquitous Computing*, Nottingham, UK, 2004, pp. 88–106.
- [31] F. Gustafsson, Particle filter theory and practice with positioning applications, *IEEE Aerosp. Electron. Syst. Mag.*, vol. 25, no. 7, pp. 53–82, 2010.
- [32] G. D. Qi, G. Pan, S. J. Li, Z. H. Wu, D. Q. Zhang, L. Sun, and L. T. Yang, How long a passenger waits for a vacant taxi—large-scale taxi trace mining for smart cities, in *2013 IEEE Int. Conf. Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing*, Beijing, China, 2013, pp. 1029–1036.
- [33] F. Z. Zhang, N. J. Yuan, D. Wilkie, Y. Zheng, and X. Xie, Sensing the pulse of urban refueling behavior: A perspective from taxi mobility, *ACM Trans. Intell. Syst. Technol.*, vol. 6, no. 3, p. 37, 2015.
- [34] D. Q. Zhang, L. Sun, B. Li, C. Chen, G. Pan, S. J. Li, and Z. H. Wu, Understanding taxi service strategies from taxi GPS traces, *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 1, pp. 123–135, 2015.
- [35] H. J. Wang, Z. H. Li, and W. C. Lee, PGT: Measuring mobility relationship using personal, global and temporal factors, in *2014 IEEE Int. Conf. Data Mining*, Shenzhen, China, 2014, pp. 570–579.
- [36] J. Hershberger and J. Snoeyink, Speeding up the Douglas-Peucker line-simplification algorithm, Vancouver, Canada: University of British Columbia, pp. 134–143, 1992.
- [37] N. Meratnia and R. A. De By, Spatiotemporal compression techniques for moving point objects, in *Proc. 9<sup>th</sup> Int. Conf. Extending Database Technology*, Heraklion, Crete, Greece, 2004, pp. 765–782.
- [38] A. Ballatore, G. McArdle, C. Kelly, and M. Bertolotto, RecoMap: An interactive and adaptive map-based recommender, in *Proc. 2010 ACM Symposium on Applied Computing*, Sierre, Switzerland, 2010, pp. 887–891.
- [39] O. Pink and B. Hummel, A statistical approach to map matching using road network geometry, topology and vehicular motion constraints, in *Proc. 11<sup>th</sup> Int. IEEE Conf. Intelligent Transportation Systems*, Beijing, China, 2008, pp. 862–867.
- [40] P. Newson and J. Krumm, Hidden Markov map matching through noise and sparseness, in *Proc. 17<sup>th</sup> ACM Sigspatial Int. Conf. Advances in Geographic Information Systems*, Seattle, WA, USA, 2009, pp. 336–343.
- [41] Y. Zheng, L. Capra, O. Wolfson, and H. Yang, Urban computing: Concepts, methodologies, and applications, *ACM Trans. Intell. Syst. Technol.*, vol. 5, no. 3, p. 38, 2014.
- [42] S. S. Chawathe, Segment-based map matching, in *Proc. 2007 IEEE Intelligent Vehicles Symposium*, Istanbul, Turkey, 2007, pp. 1190–1197.
- [43] L. Song, D. Kotz, R. Jain, and X. N. He, Evaluating location predictors with extensive Wi-Fi mobility data, in *Proc. 23<sup>rd</sup> Annual Joint Conf. of the IEEE Computer and Communications Societies*, Hong Kong, China, 2004, pp. 1414–1424.
- [44] V. Bogorny, B. Kuijpers, and L. O. Alvares, ST-DMQL: A semantic trajectory data mining query language, *Int. J. Geogr. Inf. Sci.*, vol. 23, no. 10, pp. 1245–1276, 2009.
- [45] L. O. Alvares, V. Bogorny, B. Kuijpers, B. Moelans, J. A. F. de Macedo, and A. Tietbohl Palma, Towards semantic trajectory knowledge discovery, Technical Report, Hasselt University, Belgium, 2007.
- [46] C. Parent, S. Spaccapietra, C. Renso, G. Andrienko, N. Andrienko, V. Bogorny, M. L. Damiani, A. Gkoulalas-Divanis, J. Macedo, N. Pelekis, et al., Semantic trajectories modeling and analysis, *ACM Comput. Surv.*, vol. 45, no. 4, p. 42, 2013.
- [47] Y. Z. Wang, N. J. Yuan, D. F. Lian, L. L. Xu, X. Xie, E. H. Chen, and Y. Rui, Regularity and conformity: Location prediction using heterogeneous mobility data, in *Proc. 21<sup>st</sup> ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, Sydney, Australia, 2015, pp. 1275–1284.

- [48] D. F. Lian, X. Xie, V. W. Zheng, N. J. Yuan, F. Z. Zhang, and E. H. Chen, CEPR: A collaborative exploration and periodically returning model for location prediction, *ACM Trans. Intell. Syst. Technol.*, vol. 6, no. 1, p. 8, 2015.
- [49] Q. Liu, S. Wu, L. Wang, and T. N. Tan, Predicting the next location: A recurrent model with spatial and temporal contexts, in *Proc. 30<sup>th</sup> AAAI Conf. Artificial Intelligence*, Phoenix, AZ, USA, 2016, pp. 194–200.
- [50] Y. X. Li, Y. Zheng, H. C. Zhang, and L. Chen, Traffic prediction in a bike-sharing system, in *Proc. 23<sup>rd</sup> Sigspatial Int. Conf. Advances in Geographic Information Systems*, Seattle, WA, USA, 2015, p. 33.
- [51] D. Ashbrook and T. Starner, Using GPS to learn significant locations and predict movement across multiple users, *Personal and Ubiquitous Computing*, vol. 7, no. 5, pp. 275–286, 2003.
- [52] E. Cho, S. A. Myers, and J. Leskovec, Friendship and mobility: User movement in location-based social networks, in *Proc. 17<sup>th</sup> ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, San Diego, CA, USA, 2011, pp. 1082–1090.
- [53] F. Giannotti, M. Nanni, F. Pinelli, and D. Pedreschi, Trajectory pattern mining, in *Proc. 13<sup>th</sup> ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, San Jose, CA, USA, 2007, pp. 330–339.
- [54] T. Kim, Y. S. Yue, S. Taylor, and I. Matthews, A decision tree framework for spatiotemporal sequence prediction, in *Proc. 21<sup>th</sup> ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, Sydney, Australia, 2015, pp. 577–586.
- [55] Z. H. Li, J. W. Han, M. Ji, L. A. Tang, Y. T. Yu, B. L. Ding, J. G. Lee, and R. Kays, MoveMine: Mining moving object data for discovery of animal movement patterns, *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 4, p. 37, 2011.
- [56] A. Monreale, F. Pinelli, R. Trasarti, and F. Giannotti, WhereNext: A location predictor on trajectory pattern mining, in *Proc. 15<sup>th</sup> ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, Paris, France, 2009, pp. 637–646.
- [57] S. Lee, J. Lim, J. Park, and K. Kim, Next place prediction based on spatiotemporal pattern mining of mobile device logs, *Sensors*, vol. 16, no. 2, p. 145, 2016.
- [58] P. Bhargava, T. Phan, J. Y. Zhou, and J. H. Lee, Who, what, when, and where: Multi-dimensional collaborative recommendations using tensor factorization on sparse user-generated data, in *Proc. 24<sup>th</sup> Int. Conf. World Wide Web*, Florence, Italy, 2015, pp. 130–140.
- [59] H. J. Gao, J. L. Tang, and H. Liu, gSCorr: Modeling geo-social correlations for new check-ins on location-based social networks, in *Proc. 21<sup>st</sup> ACM Int. Conf. Information and Knowledge Management*, Maui, HI, USA, 2012, pp. 1582–1586.
- [60] H. J. Guo, J. L. Tang, and H. Liu, Exploring social-historical ties on location-based social networks, in *Proc. 6<sup>th</sup> Int. AAAI Conf. Weblogs and Social Media*, Dublin, Ireland, 2012.
- [61] Y. W. Teh, A hierarchical Bayesian language model based on Pitman-Yor processes, in *ACL-44 Proceedings of the 21<sup>st</sup> International Conference on Computational Linguistics and the 44<sup>th</sup> Annual Meeting of the Association for Computational Linguistics*, 2006, pp. 985–882.
- [62] A. Sadilek, H. Kautz, and J. P. Bigham, Finding your friends and following them to where you are, in *Proc. 5<sup>th</sup> ACM International Conference on Web Search and Data Mining*, Seattle, WA, USA, 2012, pp. 723–732.
- [63] N. Du, H. J. Dai, R. Trivedi, U. Upadhyay, M. Gomez-Rodriguez, and L. Song, Recurrent marked temporal point processes: Embedding event history to vector, in *Proc. 22<sup>nd</sup> ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, San Francisco, CA, USA, 2016, pp. 1555–1564.
- [64] A. Zarezade, S. Jafarzadeh, and H. R. Rabiee, Spatio-temporal modeling of check-ins in location-based social networks, arXiv preprint arXiv: 1611.07710, 2016.
- [65] A. Noulas, S. Scellato, N. Lathia, and C. Mascolo, Mining user mobility features for next place prediction in location-based services, in *Proc. 12<sup>th</sup> Int. Conf. Data Mining*, Brussels, Belgium, 2012, pp. 1038–1043.
- [66] G. E. Hinton, S. Osindero, and Y. W. Teh, A fast learning algorithm for deep belief nets, *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.
- [67] J. B. Zhang, Y. Zheng, and D. K. Qi, Deep spatio-temporal residual networks for citywide crowd flows prediction, in *Proc. 21<sup>st</sup> AAAI Conf. Artificial Intelligence*, San Francisco, CA, USA, 2016.
- [68] J. B. Zhang, Y. Zheng, D. K. Qi, R. Y. Li, and X. W. Yi, DNN-based prediction model for spatio-temporal data, <https://www.microsoft.com/en-us/research/publication/dnn-based-prediction-model-spatial-temporal-data/>, 2017.
- [69] A. Noulas, S. Scellato, C. Mascolo, and M. Pontil, An empirical study of geographic user activity patterns in foursquare, in *Proc. 5<sup>th</sup> Int. AAAI Conf. Weblogs and Social Media*, Barcelona, Spain, 2011.
- [70] O. Bouidghaghen, L. Tamine, and M. Boughanem, Personalizing mobile web search for location sensitive queries, in *Proc. IEEE 12<sup>th</sup> Int. Conf. Mobile Data Management*, Lulea, Sweden, 2011, pp. 110–118.
- [71] A. Y. Xue, R. Zhang, Y. Zheng, X. Xie, J. Huang, and Z. H. Xu, Destination prediction by sub-trajectory synthesis and privacy protection against such prediction, in *Proc. 29<sup>th</sup> Int. Conf. Data Engineering*, Brisbane, Australia, 2013, pp. 254–265.
- [72] J. J. C. Ying, W. C. Lee, and V. S. Tseng, Mining geographic-temporal-semantic patterns in trajectories for location prediction, *ACM Trans. Intell. Syst. Technol.*, vol. 5, no. 1, p. 2, 2013.
- [73] J. Bao, Y. Zheng, D. Wilkie, and M. Mokbel, Recommendations in location-based social networks: A survey, *Geoinformatica*, vol. 19, no. 3, pp. 525–565, 2015.

- [74] A. Noulas, S. Scellato, C. Mascolo, and M. Pontil, An empirical study of geographic user activity patterns in foursquare, in *Proc. 5<sup>th</sup> International AAAI Conference on Weblogs and Social Media*, Barcelona, Spain, 2011.
- [75] J. Bao, Y. Zheng, and M. F. Mokbel, Location-based and preference-aware recommendation using sparse geo-social networking data, in *Proc. 20<sup>th</sup> Int. Conf. Advances in Geographic Information Systems*, Redondo Beach, CA, USA, 2012, pp. 199–208.
- [76] Y. Zheng, Q. N. Li, Y. K. Chen, X. Xie, and W. Y. Ma, Understanding mobility based on GPS data, in *Proc. 10<sup>th</sup> Int. Conf. Ubiquitous Computing*, Seoul, Korea, 2008, pp. 312–321.
- [77] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge, UK: Cambridge University Press, 2008.
- [78] Y. Arase, X. Xie, T. Hara, and S. Nishio, Mining people’s trips from large scale geo-tagged photos, in *Proc. 18<sup>th</sup> ACM Int. Conf. Multimedia*, Firenze, Italy, 2010, pp. 133–142.
- [79] Y. Wei, Y. Zheng, and Q. Yang, Transfer knowledge between cities, in *Proc. 22<sup>nd</sup> ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, San Francisco, CA, USA, 2016, pp. 1905–1914.
- [80] J. Y. Zhu, C. Zhang, H. C. Zhang, S. Zhi, V. O. K. Li, J. W. Han, and Y. Zheng, pg-causality: Identifying spatiotemporal causal pathways for air pollutants with urban big data, *IEEE Trans. Big Data*, doi: 10.1109/TBDA.2017.2723899.
- [81] Y. Liu, Y. Zheng, Y. X. Liang, S. M. Liu, and D. S. Rosenblum, Urban water quality prediction based on multi-task multi-view learning, in *Proc. 25<sup>th</sup> Int. Joint Conf. Artificial Intelligence*, New York, NY, USA, 2016.
- [82] H. J. Wang, D. Kifer, C. Graif, and Z. H. Li, Crime rate inference with big data, in *Proc. 22<sup>nd</sup> ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, San Francisco, CA, USA, 2016, pp. 635–644.
- [83] C. Lopez and Y. P. P. Chen, Using object and trajectory analysis to facilitate indexing and retrieval of video, *Knowl.-Based Syst.*, vol. 19, no. 8, pp. 639–646, 2006.
- [84] Z. Y. Cheng, J. Caverlee, K. Lee, and D. Z. Sui, Exploring millions of footprints in location sharing services, in *Proc. 5<sup>th</sup> Int. AAAI Conf. Weblogs and Social Media*, Barcelona, Spain, 2011.
- [85] L. Backstrom, E. Sun, and C. Marlow, Find me if you can: Improving geographical prediction with social and spatial proximity, in *Proc. 19<sup>th</sup> Int. Conf. World Wide Web*, Raleigh, NC, USA, 2010, pp. 61–70.



**Ruizhi Wu** received the BS degree from Hangzhou Dianzi University in 2012. He is currently pursuing the PhD degree in the School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China. His research interests include spatial-temporal data mining, data mining, and machine learning.



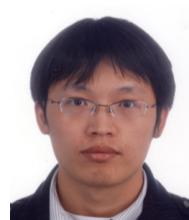
**Guangchun Luo** received the BS, MS, and PhD degrees in computer science from University of Electronic Science and Technology of China, Chengdu, China, in 1995, 1999, and 2004, respectively. He is currently a professor of computer science at the University of Electronic Science and Technology of China, Chengdu, China. His research interests include computer networking, cloud computing, and big data.



**Chengzong Peng** received the BS degree from University of California, Irvine (UCI) in 2017. He is currently working as a research assistant in School of Computer Science, UESTC. His research interests include database and big data technologies.



**Ling Tian** received the BS, MS, and PhD degrees in computer science from University of Electronic Science and Technology of China, Chengdu, China, in 2003, 2006 and 2010, respectively. She is currently an associate professor with the School of Computer Science and Engineering, University of Electronic Science and Technology of China. She was a visiting scholar in Georgia State University in 2013. Her research interests include digital multimedia, cloud computing, and big data.



**Junming Shao** received the PhD degree with highest honors (Summa Cum Laude) from the University of Munich, Germany, in 2011. He became the Alexander von Humboldt Fellow in 2012. Currently, he is the professor of Computer Science at the University of Electronic Science and Technology of China. His research interests include data mining and neuroimaging. He not only published papers on top-level data mining conferences like KDD, ICDM, SDM (two of those papers have won the Best Paper Award), but also published data mining-related interdisciplinary work in leading journals including *Brain, Neurobiology of Aging*, and *Water Research*.