IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

SPECIAL SECTION ON THEORETICAL FOUNDATIONS FOR BIG DATA APPLICATIONS: CHALLENGES AND OPPORTUNITIES

Received February 29, 2016, accepted March 27, 2016, date of publication April 13, 2016, date of current version May 23, 2016.

Digital Object Identifier 10.1109/ACCESS.2016.2553681

A Survey on Trajectory Data Mining: Techniques and Applications

ZHENNI FENG[1] AND YANMIN ZHU[1,2], (Member, IEEE)
[1]Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
[2]Shanghai Key Laboratory of Scalable Computing and Systems, Shanghai Jiao Tong University, Shanghai 200240, China

Corresponding author: Y. Zhu (yzhu@sjtu.edu.cn)

This work was supported in part by the 973 Program under Grant 2014CB340303, in part by the National Natural Science Foundation of China under Grant 61472254, Grant 61170238, and Grant 61420106010, in part by the Science and Technology Commission of Shanghai under Grant 14511107500, in part by the National Research Foundation Singapore Energy and Environmental Sustainability Solutions for Megacities under the Campus for Research Excellence and Technological Enterprise framework, and in part by the Program for New Century Excellent Talents in University and the Program for Changjiang Scholars and Innovative Research Team in University (Young Scholars), China.

ABSTRACT Rapid advance of location acquisition technologies boosts the generation of trajectory data, which track the traces of moving objects. A trajectory is typically represented by a sequence of timestamped geographical locations. A wide spectrum of applications can benefit from the trajectory data mining. Bringing unprecedented opportunities, large-scale trajectory data also pose great challenges. In this paper, we survey various applications of trajectory data mining, e.g., path discovery, location prediction, movement behavior analysis, and so on. Furthermore, this paper reviews an extensive collection of existing trajectory data mining techniques and discusses them in a framework of trajectory data mining. This framework and the survey can be used as a guideline for designing future trajectory data mining solutions.

INDEX TERMS Trajectory data mining, big data applications, data mining techniques.

I. INTRODUCTION

Nowadays, there have been many technologies which provide positioning services e.g., Global Position Systems (GPS), Radio Frequency Identification (RFID), location estimation of 802.11, smart phone sensors, GSM beacons, infrared or ultrasonic systems and so on [1]. As a consequence, it is becoming easier to generate large-scale trajectory data of tracking traces of moving objects.

A trace of a moving object in geographical space is continuous while a trajectory is only a sample of location points that the moving object passes as shown in Fig. 1. Typically, a spatial trajectory, as a simplest case of trajectory data, is represented by a sequence of timestamped locations, e.g., $\langle (p_0, t_0), (p_1, t_1), \cdots, (p_7, t_7) \rangle$ in Fig. 1. Duration and sampling rate of a trajectory depend on applications.

Trajectory data are collected from various sources. One of the most common types is generated by GPS-equipped vehicles. Besides, other kinds of trajectories probably come from smart phones, online check-in data, geo-tagged messages or media in social networks, RFID readers, and so on. Consequently, moving objects can be human beings, animals, vehicles, and even natural phenomena (e.g., hurricanes).

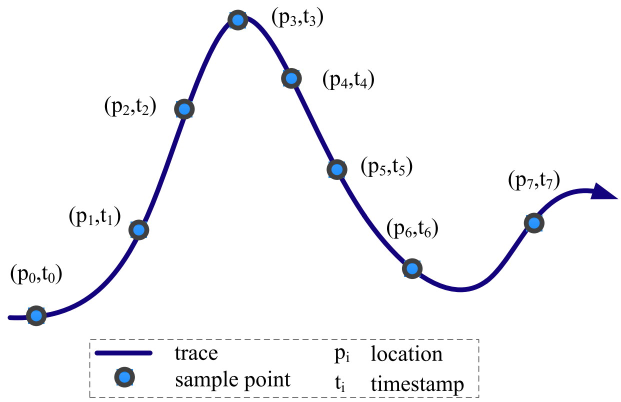


FIGURE 1. A trajectory is generated by sampling from a continuous trace.

There exist a wide spectrum of applications driven and improved by trajectory data mining, such as path discovery, location/destination prediction, movement behavior analysis for individual or a group of moving objects, making sense of trajectories and other applications of urban service. These applications significantly benefit the common people, commercial organizations and government agencies.

However, it is challenging to manage, process and mine trajectory data [2], [3]. We take several challenges as example. *Firstly*, it is a nontrivial task to store a huge volume of trajectory data which are rapidly accumulated. *Secondly*, it is intractable to define a similarity metric for comparing trajectories (which is a fundamental functionality in trajectory data mining) since trajectories are probably generated with different sampling strategies or at different sampling rates. *Thirdly*, processing queries on the vast amount of trajectory data is highly difficult in terms of space or time complexity.

To address these issues, an extensive collection of approaches have been proposed and we classify them according to the main procedure of trajectory data mining. Furthermore, we propose a framework that reorganizes these approaches and then provide a comprehensive survey on trajectory data mining. Roughly speaking, there are three layers in the framework, i.e., *data collection*, *trajectory data mining techniques*, *applications*.

Specifically, the layer of trajectory data mining techniques contains five components listed as follows:

- **Preprocessing:** In the preprocessing phase, trajectories are usually cleaned, segmented, calibrated, sampled for representatives, or inferred from uncertain trajectories.
- **Data management:** Sometimes, trajectories are compressed or simplified before being stored. Besides, efficient or scalable storage systems are supposed to be built. Furthermore, appropriate index structures are also necessary to support query processing.
- **Query processing:** There are various queries that have to be processed to retrieve data, e.g., location-based queries, range queries, nearest neighbor queries, top-k queries, pattern queries, aggregate queries and other application-specific queries. These queries are processed based on an underlying storage system and index structure.
- **Trajectory data mining tasks:** Trajectory data mining tasks are summarized and classified into several categories, i.e., pattern mining, clustering, classification and knowledge discovery.
- **Privacy protection:** Privacy-preserving is a crucial problem in every procedure of trajectory data mining. Several examples are provided to illustrate how to process trajectory data as well as to protect sensitive information of users.

The rest of the paper is structured as follows. Section II offers some definitions. A framework characterizing the whole procedure of trajectory data mining is presented in Section III. The next five sections, Section IV to Section VIII, cover the important components of trajectory data mining techniques. Next, applications of trajectory data mining are discussed in Section IX. Section X discusses a few open issues. Finally, the paper is concluded in Section XI.

II. DEFINITIONS

In the section, we define some primary terminologies, e.g., trajectory, semantic trajectory, road network, path.

Definition 1 (Trajectory): A trajectory of a moving object is a discrete trace that the moving object travels in geographical space. Generally, it is a sequence of geo-locations with corresponding timestamps in spatio-temporal space, i.e., $\mathcal{T} = \{< p_1, t_1 >, < p_2, t_2 >, \cdots, < p_n, t_n >\}$, where each element $< p_i, t_i >$ indicates a moving object is at location p_i at timestamp t_i. Further, elements are sorted by timestamps, i.e., $t_j < t_k$ if $1 \leq j < k \leq n$.

A moving object can be a person, an animal, a vehicle, a mobile device, or even a phenomenon. A trajectory of a person records one's trace for a period of time. For example, a trajectory of a person in a daytime may record his path to work in the morning and his path to home at night. A trajectory of an animal describes its trace generated by daily activities such as running. A trajectory of a vehicle is recorded by a GPS device installed in the vehicle and usually reports locations of the vehicle at a fixed rate, e.g., every second or every minute.

A location is usually expressed by a tuple of $\langle$*longitude*, *latitude*$\rangle$ which is recorded by a GPS device. Each tuple of $\langle$*longitude*, *latitude*$\rangle$ corresponds to a unique point in geographical space. A special kind of trajectory data is RFID data. There are two types of devices in RFID technology, i.e., tags (which emit radio signals with identification information) and readers (which detect signals from tags). Generally, a moving object is a tag device, i.e., a good in a warehouse. A location is expressed by identification of a reader which detects signal from that tag. Essentially, locations of a moving object are recorded by corresponding geographical areas of readers that detect its signals.

Sampling rates of trajectory data vary greatly from data source to data source. Due to energy and storage limitation, trajectories of different kinds of moving objects are sampled at different rates. Trajectory data of vehicles usually have higher sampling rates than those of mobile devices since vehicles can provide adequate battery and storage.

Specifically, a trajectory introduced above can also be called a *geographical trajectory* or a *raw trajectory*. Then, we introduce a special kind of trajectory data that integrate geographical positions with semantic meaning.

Definition 2 (Semantic Trajectory): A trajectory is called a semantic trajectory when its locations are associated with semantic entities.

Semantic trajectories are often generated by tagging a location point with a meaningful place in real world apart from numerical coordinates, such as check-in data. Besides, when a geographical trajectory is associated with description text which expresses one's feeling and emotion, it is also a kind of semantic trajectory data.

When considering trajectories generated by vehicles, we often refer to a road network and paths in the road network. Their definitions are given as follows.

Definition 3 (Road Network): A road network is a directed graph, $G = (V, E)$, where V and E are a vertex set and an edge set, respectively. A vertex, $v_j \in V$, is a road junction or road end. An edge, $e_k = v_p v_q \in E$, denotes a directed road segment, on which travel direction of moving objects is from v_p to v_q.

Definition 4 (Path): A path, $\mathcal{P} = \langle e_1, e_2, \cdots, e_{|\mathcal{P}|} \rangle$, represents a sequence of edges in a road network, where $e_i \in E$ and $e_i \neq e_j$ if $i \neq j$. Specifically, consecutive edges e.g., e_i and e_{i+1}, must share a vertex which is end vertex of e_i and source vertex of e_{i+1}.

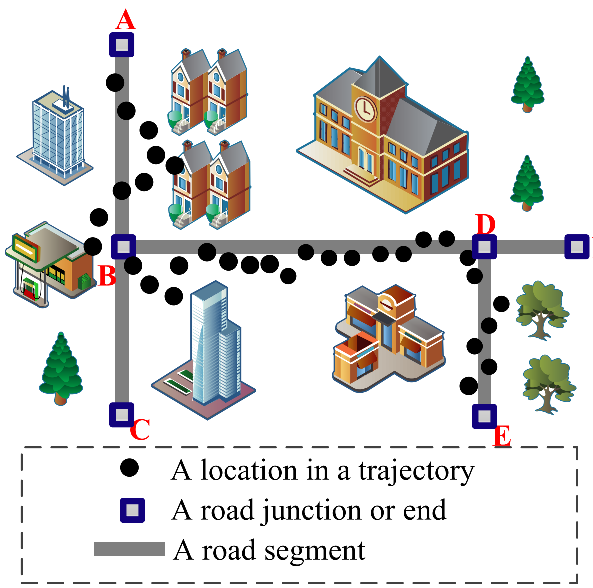


FIGURE 2. **An example of a road network and paths. All location points in a trajectory are matched to edges in a road network and as a result paths are generated. Suppose a road network is denoted by $G = (V, E)$, where its set of vertices is $V = \{A, B, C, D, E, F\}$ and its set of directed edges is $E = \{AB, BA, BC, CB, BD, DB, DE, ED, DF, FD\}$. The generated path after map-matching is $\langle AB, BD, DE \rangle$.**

Fig. 2 gives an example of a road network and paths. Each location point in the collected trajectory is first matched to a specific road segment on the map. The path that the moving object travels is $\langle AB, BD, DE \rangle$.

III. A FRAMEWORK OF TRAJECTORY DATA MINING

In the section, we propose a framework that summarizes a whole procedure of trajectory data mining as shown in Fig. 3. It is worth noting that not every step in the layer of trajectory data mining techniques is necessary and it depends on requirement of applications and collected data.

Firstly, trajectory data are generated by various moving objects and collected from multiple data sources. In the paper, we omit details in data collection. *Then*, main part of trajectory mining techniques are presented with five components, i.e., preprocessing, data management, query processing, trajectory data mining tasks, and privacy protection. *Finally*, in the layer of applications, we review an extensive of applications from six categories.

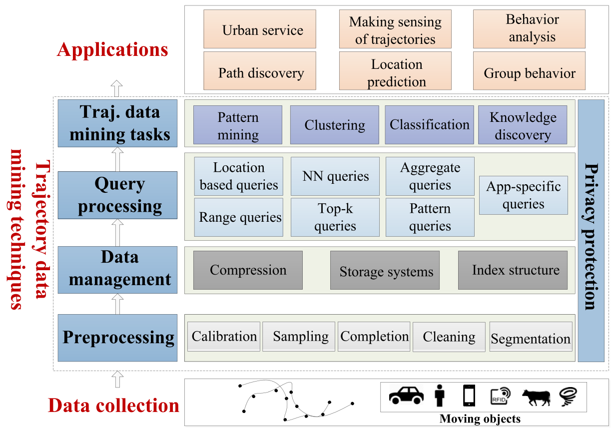


FIGURE 3. **A framework of trajectory data mining.**

The layer of trajectory data mining techniques is organized as follows. Preprocessing attempts to improve quality of trajectory data and to partition trajectories into sub-trajectories for further processing. Data management solves the problem of storing a huge amount of trajectory data in an efficient and scalable manner. Query processing aims to retrieve appropriate data from the underlying storage system efficiently. The component of trajectory data mining tasks summarizes several important types of mining tasks. Protecting privacy of users with privacy-preserving techniques is an essential problem throughout these four components above and thus it can be combined with any component.

All components of trajectory data mining techniques and the layer of applications will be discussed successively in following sections.

IV. PREPROCESSING

Preprocessing is a basic step that performs at the beginning and it aims at improving quality of trajectory data and generating sub-trajectories. In the section, we presented five common operations in preprocessing phase.

A. CLEANING

Outliers existing in trajectory data highly degrade performance of trajectory data mining techniques. Previous work [4] trys to detect suspicious moving objects or to capture features of many abnormal moving trajectories.

Due to ambiguity of RFID data, i.e., there existing no deterministic location given multiple readers detected an object, cleaning trajectory data aims to discard impossible locations or trajectories exploiting specific constraints, e.g., maximum speed, unreachability constraints [5].

B. SEGMENTATION

In many application scenarios, a trajectory is partitioned into sub-trajectories [6], each of which is often called a segment, a partition or a frame. Generating sub-trajectories is rational because it corresponds to underlying structures in trajectory data, e.g., a path with multiple road segments.

A partition-and-summarization approach [7] that attempts to generate human-readable description of trajectory data also divides a trajectory into several partitions according to behaviors of moving objects.

Trajectories are partitioned into frames [8], [9] in order to efficiently store sample points of a moving object which are aligned by time intervals, leveraging state-of-the-art column-oriented storage system.

C. COMPLETION

Due to consideration of storage and transmission, trajectories are mostly collected at relatively low sampling rates, only providing partial observations of actual routes. These trajectories are called *uncertain trajectories*. It is a crucial problem to infer uncertain trajectories with the aid of various constraints in reality. Existing studies [10]–[14] aim to complete these trajectories and support trajectory data mining tasks.

D. CALIBRATION

Trajectories that represent discrete approximation of original routes with different sampling strategies and different sampling rates are heterogeneous. Heterogeneity has a negative effect on measurement of trajectory similarity, e.g., it is difficult to compare two trajectories derived with different sampling strategies by directly utilizing spatial proximity based similarity measures like Euclidean distance. Su *et al.* [15], [16] focus on transforming such heterogeneous trajectories to ones with unified sampling strategies.

E. SAMPLING

In the field of trajectory data mining, approaches are often operated on a large trajectory databases and thus operations are complicated, expensive and time-consuming. Trajectory sampling approaches [6], [17], [18] aim to reduce a large trajectory database appropriately, taking only the most representative samples of original trajectory database. Certainly, the subset of samples should encapsulates mobility patterns hidden in the original trajectory database.

Pelekis *et al.* [6] represent a trajectory as a symbolic vector that quantifies the representativeness of each trajectory, and then propose an unsupervised method to sample representative trajectories. In another work [17], trajectory sampling are operated on sub-trajectories, i.e., segments of trajectories, deriving local trajectory descriptors that represent line segments. A recent study [18] whose objective is supporting trajectory aggregate queries also addresses the trajectory sampling problem. It focuses on approximating queries processing with response-time constraints.

V. DATA MANAGEMENT

How to store a huge amount of trajectory data is a fundamental problem in trajectory data mining. In the section, we discuss the problem from two aspects.

A. COMPRESSION

It is relatively problematic to store or transmit a huge amount of trajectory data created by location-acquisition techniques. As some location points in a trajectory are often redundant, trajectory compression algorithms [19]–[21] are promising to reduce storage requirements and communication loads. A compression approach oftem makes a tradeoff between compression ratio and maximum error. Generally, the higher the compression ratio, the poorer the quality of compressed trajectory data. Typical approaches are line generation and delta compression.

Trajic [19] achieves a relatively good compression ratio if users are willing to tolerate a small amount of error. It contains a predictor which predicts next data point, and a residual encoding scheme that generates small residuals to compensate difference between predicted value and actual value.

PRESS [20] is a novel framework that separates spatial representation from temporal representation and proposes a spatial compression algorithm and a temporal compression algorithm, respectively. Thus, compression is efficient due to parallelism of spatial and temporal compression procedures. Spatial-temporal queries can be executed without fully decompressing trajectory data. It is also a lossy compression approach like Trajic.

Liu *et al.* [21] propose an online error-bounded compression system. The approach creates a virtual coordinate system which is centered at a starting point and builds convex-hulls to bound points.

Some other studies claim that original trajectory data are significantly large and are suggested to be simplified [22], [23]. Such simplification can be regarded as a kind of lossy trajectory compression.

B. STORAGE SYSTEMS & INDEX STRUCTURES

How to store the tremendous amount of trajectory data produced by location-acquisition technologies is a crucial problem [8], [9], [24]–[27].

To boost performance of query processing for trajectory data, Wang *et al.* [8], [9] propose SharkDB, a novel in-memory based trajectory storage system. To employ column-oriented storage, trajectories are partitioned into frames, and then frames are further compressed and well-structured to better support trajectory data mining.

TrajStore [24] is a dynamic storage system which is able to retrieve all data in a particular region efficiently. Instead of simply indexing geo-spatial data, TrajStore slices trajectories into sub-trajectories according to spatial-temporal regions and stores packed data in each region together.

An index structure called TrajTree [25] is developed to manage trajectory data especially for retrieval tasks like k-NN queries. Popa *et al.* [26] propose another kind of index structure that is suitable for trajectory data flows and attempts to achieve an optimal retrieval cost of spatial-temporal queries. Another study by Ni and Ravishankar [27] present a parametric space indexing method that uses polynomial approximations to index line segments in trajectory data.

VI. QUERY PROCESSING

Retrieving data from an underlying storage system is a crucial operation. The objective of retrieval is to find appropriate data efficiently. In the section, we summarize various kinds of queries that are designed to retrieve trajectory data.

A. LOCATION-BASED QUERIES

A location-based query attempts to find trajectories that are close to all query locations where the query is a small set of locations with or without a specific order constraint. One typical application is route recommendation for a trip to multiple places. For example, in Fig. 4, a traveler at the source location S wants to find a suitable path that passes query locations A and B to the destination Chen *et al.* [28] explore this kind of problem and find the k best-connected trajectories (k-BCT) that each of which connects all query locations.

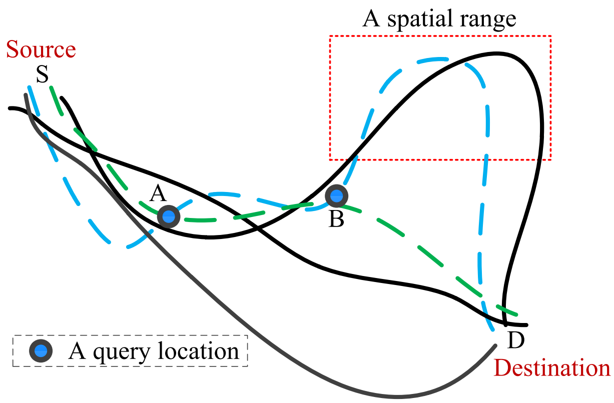


FIGURE 4. **A location based query is possibly to find trajectories that successively pass *A* and *B*. Two dotted trajectories are possible candidate trajectories. A spatial range in a range query is denoted by a dashed rectangle in red.**

Yan *et al.* [29] also address an location-based searching problem where importance of location points is differentiated. Specifically, a location with a geo-tagged photograph is more important than other locations without semantics. The proposed query is to find the k most important connected trajectories (k-ICT). Besides, efficient algorithms are proposed to answer k-ICT queries, leveraging both spatial proximity and temporal duration at each important place.

B. RANGE QUERIES

Range queries on trajectory are essential in a wide spectrum of trajectory data mining applications. A query that specifies a value to fall in a lower and upper boundary is regarded as range query, e.g., finding all trajectories of a specific traveler between 11 am and 2 pm. In Fig 4, a spatial range query is probably to search all trajectories that pass a spatial region denoted by a dashd rectangle.

Some studies [30], [31] investigate this kind of queries for uncertain trajectories. Zheng *et al.* [30] first create a probabilistic model to represent possible locations of a moving object at a specific time, and design an effective index structure to process probabilistic range queries. In another work [31], a range query means to retrieve trajectories, each of which is consistently covered by a given area with a high probability during a certain period of time.

C. NEAREST NEIGHBOR (NN) QUERIES

Nearest neighbor search is another fundamental query in spatial-temporal trajectory data mining [32], [33]. Güting *et al.* [32] address a problem of finding the k nearest neighbors to a given trajectory in the trajectory database for a specific time interval.

Another work [33] extends nearest neighbors queries to a more realistic case, the probabilistic nearest neighbor queries, where trajectories are uncertain. Based on representation of uncertain trajectories as stochastic processes, several kinds of probabilistic nearest neighbor queries with inputs of a given trajectory and a time interval are investigated.

D. TOP-*k* QUERIES

KSQ [34] focuses on finding the k most similar trajectories for a given trajectory on uncertain trajectories. The key to a rational solution is to appropriately quantify the (dis)similarity of two uncertain trajectories. A novel distance metric is proposed and a scalable index structure is further built to support trajectory data mining tasks.

E. PATTERN QUERIES

Vieira *et al.* [35] design a query with a given pattern, i.e., selecting trajectories based on a specific motion pattern. Specifically, this query typically offers a set of predicates or constraints that should be satisfied in a specific order. One predicate or constraint can be a range condition as well as a nearest-neighbor condition. The crucial problem in processing this kind of queries is to create a powerful description that represents patterns in the query as regular expressions.

F. AGGREGATE QUERIES

Li *et al.* [18] introduce another kind of queries called trajectory aggregate queries whose querying results are not trajectories in a trajectory archive but an aggregated measurement. This kind of query, as a kind of function, attempts to retrieve statistics of trajectories that traverse a given region in a specific time interval. An example of an aggregate query for vehicular trajectory data is to retrieve average velocity passing a user-specific road segment.

G. OTHER APPLICATION-SPECIFIC QUERIES

There is a wide spectrum of other application-specific queries and in the following we give a few examples.

One study [36] explores *keyword queries for semantic trajectories*. It is very useful in a lot of location-based applications, e.g., intelligent tourist guide and trip planning. For instance, a tourist would like to have a dinner in a *restaurant* and to watch a film in a *theatre*. Consequently, two keywords in italic are extracted to search reference trajectories. The biggest difference between this kind of queries and conventional spatial queries is that there exist no geo-locations in this kind of queries. The work aims to support efficient trajectory search of approximate keywords in semantic trajectory data. Approximate keyword search is a realistic extension of exact keyword search in cases of miss-spelling or fuzzy search conditions.

Zheng *et al.* [37] introduce a special case of semantic trajectories, called activity trajectories, and discuss a problem of efficient search for activity trajectories. In activity trajectories, semantic meanings that are attached to locations is information about user activities at particular places, e.g., sport, shopping, dining and entertaining. This work aims to identifying trajectories with similar activities to a given trajectory. It is also of great importance to route recommendation and trip planning.

In contrast to conventional location-based queries in spatial domain only, Shang *et al.* [38] take both spatial domain and textual domain into consideration in a query, i.e., input information in a query contains a set of locations specified by users as well as a collection of textual attributes that explain users' preferences. In trip recommendation applications, textual attributes can be, for example, a budget constraint.

Another work by Shang *et al.* [39] introduces another kind of queries that consider personalized preference of individual users. In other words, a query is specified in spatial terms as well as user-specific significance for each sampling point in trajectory data. In contrast to equal weight for every sample point in trajectory data, the work allows users to differentiate weights according to their preferences.

VII. TRAJECTORY DATA MINING TASKS

Trajectory data mining tasks are classified into several categories according to type of each task.

A. PATTERN MINING

Pattern mining is to analyze mobility patterns of a moving object or multiple moving objects together. There are various types of patterns, such as gathering/group patterns [40]–[42], sequential patterns [43] and periodic patterns [44], [45].

Regarding each trajectory as a sequence, a sequential pattern is often defined as a subsequence that at least δ trajectories share the subsequence, where δ is a user-specific threshold. Zheng *et al.* [41] address a problem of mining sequential patterns in semantic trajectories, leveraging a two-step procedure, SPLITTER, to discover fined-grained sequential patterns. SPLITTER first retrieves a collection of coarse patterns by grouping similar places together and then derives fine-grained patterns by splitting a coarse pattern in a top-down manner.

A periodic pattern is another common trajectory pattern which is significant to understand behavior of moving objects. Li *et al.* [44], [45] address a problem of mining periodicity. The work solves two crucial sub-problems of detecting periods and mining periodic movement behavior based on reference locations and probabilistic models, respectively.

B. CLUSTERING

A branch of research [46]–[48] considers a problem of trajectory clustering. It is useful to cluster trajectories into groups with similar movement patterns.

TODMIS [47] is a general framework for mining communities from multiple sources of trajectories. Groups of moving objects are identified based on trajectory-related information (e.g., spatial dispersion, temporal duration, movement velocity) as well as semantic meaning of locations.

Liu *et al.* [48] address a problem of identifying hot spots of moving vehicles, which is essentially a clustering problem. Since hot spots can be interpreted as areas of high crowdness of vehicles, clustering is a promising method to solve the problem. In contrast to conventional density-based clustering, the work employs a mobility-based clustering, whose rationale is a simple observation that a vehicle of high mobility (speed) probably implies a low crowdness and vice versa. The mobility-based clustering is less sensitive than the density-based clustering to the size of trajectory dataset.

C. CLASSIFICATION

Classification of trajectories is to build a model on training data and then to apply the trained model to predict the labels of test trajectories. Patel *et al.* [49] focus on a classification problem and introduce duration information to boost prediction accuracy. The method incorporates not only spatial information, e.g., spatial distribution, shapes of trajectories but also duration information as features for classification, as duration information greatly contributes to differentiating moving objects that travel at different velocities.

D. KNOWLEDGE DISCOVERY

Besides, sometimes we can benefit from trajectory data from knowledge that discovered in trajectory data mining. Yuan *et al.* [50], [51] focus on discovering regions of different functions in a city. The knowledge can help citizens to make decisions, e.g., whether to invest in real estate. Another two studies [52], [53] pay attention to an event detection problem from different aspects. Detected events are other kinds of valuable knowledge.

VIII. PRIVACY PROTECTION

There exist many privacy-preserving solutions [54]–[57] in trajectory data mining techniques. It is a challenging problem to support trajectory data mining as well as protecting privacy of users.

Andrienko *et al.* [54] employ a privacy-respectful manner that transforms original geo-referenced data to trajectories in an abstract semantic space upon which trajectory data are processed further.

The objective of SmartTrace [55] is to find the most similar trajectories with a given query trajectory. The method attempts to perform a distributed similarity measure where users are not necessary to upload their data and thus protects their sensitive geo-locations.

Another work by Pelekis *et al.* [56] facilitates privacy-aware sharing of mobility data and develops a trajectory engine to provide restricted access to trajectory database.

Kong *et al.* [57] introduce a homomorphic encryption scheme to ensure privacy-preserving for trajectory recovery in a crowdsourcing manner.

IX. APPLICATIONS OF TRAJECTORY DATA MINING

A wide spectrum of applications are driven by trajectory data mining. In the section, we classify these applications into following six types. We then introduce each kind of applications through a few examples.

A. PATH DISCOVERY

Path discovery is one of the most common applications of trajectory data mining. It is extremely important to find the most suitable path in many application scenarios. Exact meaning of the word "suitable" depends on applications. It can be the fastest, the shortest, the most popular, and so on. A lot of research papers [58]–[63] in the field have been published.

Path discovery, also called route discovery, is to find at least one path that satisfy a predefined objective given a source and a destination. Routes must be derived based on a specific road network. Furthermore, geographical locations in numerical style in trajectories should be matched to a map in order to derive candidate paths or path segments. Historical trajectories on the road network provide valuable intelligence to estimate, compare and even construct candidate routes.

The fastest path problem is a modification of the shortest path problem. It can be solved by setting edge costs to be time-related factors, e.g., travel time, instead of road distances. However, sometimes the problem is generalized to multiple destinations [58]. The objective is to minimize the cost of a combination of destinations.

When planning a trip in an unfamiliar area, people usually try to find *the most frequent path* between two locations [59]. Furthermore, in a more realistic scenario [61], an problem is possible to find the most frequent path in a certain time period, i.e., given a time period T, a source v_s and a destination v_d, searching the most frequent path during T. Apart from time period constraints, Wei *et al.* [60] further consider a situation of *uncertain trajectories*, where trajectories are generated at a very low sampling rate due to multiple reasons, i.e., hardware limitations, privacy concerns, energy constraints. In Fig. 5, a trajectory is generated by sampling at a low frequency with only S, A, B, D as sample points. The movement between A and B are uncertain and there are various routes, e.g., colored lines in Fig. 5.

Generally speaking, the most frequent paths outperform the fastest paths or the shortest paths since the frequent ones

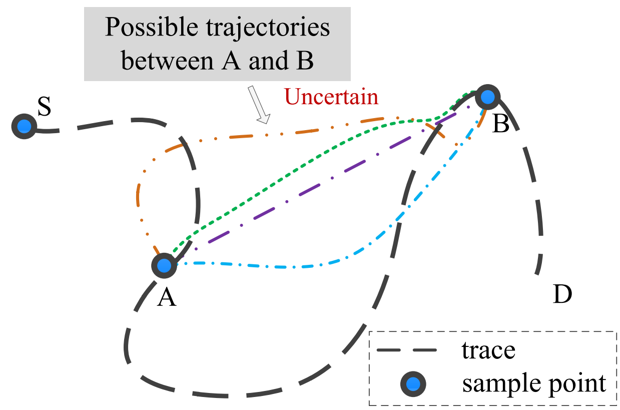


FIGURE 5. An example of uncertain trajectories. A trajectory from *S* to *D* is generated at a relatively low sampling rate and only two points *A* and *B* are sampled. Movement between *A* and *B* is uncertain.

reflect common routing preferences of previous travelers. It also helps to reduce the risk of failed paths which are possibly unpaved, dangerous or blocked by a recent road work.

In terms of public transportation, people's real demand for public transportation are employed to identify and optimize existing flawed bus routes, thus improving utilization efficiency of public transportation [62].

To take into account various driving preferences, Dai *et al.* [63] propose a recommendation system that chooses different routes for drivers with different driving preferences. This kind of *personalized route recommendation* avoids flaws of previous unique recommendation and improves quality of user satisfaction.

B. LOCATION/DESTINATION PREDICTION

Location based services (LBSs), also called location-aware services, are increasingly beneficial to people in urban areas. It has been revealed that human mobility is extraordinary regular and thus predictable. Many location based applications require location prediction or destination prediction to send advertisements to targeted consumers, to recommend tourist spots or restaurants, or to set destinations in navigation systems.

Destination prediction is closely related to path discovery. If an ongoing trip matches part of a frequent route in a dataset of historical trajectories, the destination of the frequent route is possibly the destination of the ongoing trip. However, there exist a few constraints in real world scenarios. Research examples are stated as follows.

Xue *et al.* [64], [65] point out a *data sparsity* problem, which indicates that available trajectories are too few to cover all possible trajectories. To tackle the data sparsity problem, all trajectories are decomposed into sub-trajectories, and then synthesized trajectories are generated by connecting sub-trajectories together. An expanded set of trajectories that can support destination prediction is exponentially increased by this method. In this paper, privacy protection is also considered to protect sensitive location information of users.

Noulas *et al.* [66] focus on a problem of predicting the next place that a user will visit, by exploring human mobility patterns. A large amount of check-in data are utilized to study human movement with a qualitative representation. Then a set of features which are corresponding to potential factors that may drive movement of users are extracted.

Apart from mobility patterns of individual users, another study [67] further thinks about social conformity of users, i.e., one's movement is influenced by others'. Both regularity and conformity are considered to improve the predictive power. Moreover, heterogeneous mobility datasets e.g., GPS trajectories, cellular tower data, WiFi signals, smart card transactions, check-in locations from online social networks instead of a single type of trajectories are introduced to boost prediction performance.

C. MOVEMENT BEHAVIOR ANALYSIS

Trajectory data provide a lot of opportunities to analyze movement behavior of moving objects [68]–[73]. Discovery of movement patterns is crucial for understanding human behavior. One important challenge in this topic is to extract high-level semantics of behavior, i.e., inferring underlying purposes or roles of moving objects.

Renso *et al.* [68] propose an approach to understand behavior of people who move in a geographical context by extracting mobility behavioral patterns. Then, human behavior is inferred from these patterns which are mined from trajectory data.

Predicting human behavior accurately under emergency is a crucial issue for disaster alarming, disaster management, disaster relief and societal reconstruction after disasters. Song *et al.* [69] analyze emergency behavior of human beings and their mobility patterns after a big nuclear accident in Japan, leveraging a large human mobility database. It is proved that emergency behavior after disasters sometimes correlates with their normal mobility patterns. Furthermore, several impacting factors, e.g., social relationship, intensity of a disaster, damage level, new reporting, population flow, are investigated and thus a predictive model is derived.

Another study [70] addresses a problem of detecting roles of moving objects from trajectory data. It is assumed that the intrinsic structure, i.e., the distribution of behavior, characterizes each role. Consequently, the role of a moving object can be identified by exploring structures of trajectories.

Human mobility behavior can be studied from spatial, temporal and social aspects. Gao *et al.* [71] present a comprehensive analysis of temporal effects in modeling mobility behavior. It has been studied that human mobility exhibits strong temporal cyclic patterns in the period of hour, day or week.

Liu *et al.* [72] propose a method to model trajectories in terms of user decision on visiting a point of interest (POI) and conduct rationality analysis upon trajectory behavior. Rationality of trajectory behavior is explored through several impacting factors.

Another recent work [73] explores individual human mobility patterns by studying a large number of anonymous position data from mobile phone users and reveals a high degree of temporal and spatial regularity in human trajectories.

D. GROUP BEHAVIOR ANALYSIS

Moving objects, especially people and animals, sometimes tend to form groups or clusters due to their social behavior. For instance, movement of a person is affected by not only personal activities, but also social ties with that of the groups he belongs to. Besides, a gathering pattern, as a novel modeling of trajectory patterns, describes movement pattern of a group of moving objects. Examples include celebrations, parades, traffic congestion, large-scale business promotions, protests, etc. The topic of mining gathering patterns or group patterns has attracted a lot of research attention [40]–[42], [74], [75]. Informally, a gathering in reality indicates an unusual or significant event.

Zheng *et al.* [40] introduce a gathering pattern generated by a dense and continuing group of moving objects. Gathering removes requirement for coherent membership in traditional group patterns (e.g., flock, convoy and swarm), leading to a general membership that allows moving objects to enter or leave its group anytime. An extension [41] derives an efficient online discovery approach, i.e., in an incremental manner to incorporating newly generated trajectory data.

Another study [42] also aims at efficiently discovering moving objects which move together. A group is defined as a cluster that at least m moving objects being densely connected for at least a certain duration of time. It is very different from gathering meaning aforementioned. Besides, a sampling-independent approach is proposed to avoid flaws of sampling dependent ones, e.g., convoy, swarm.

Gupta *et al.* [74] first address a problem of efficiently modeling individual and group behavior and then present a simulation framework that simulates people's movement behavior in order to generate spatio-temporal movement data. The simulation is of great significance since a large amount of movement data in public domain are limited and unavailable in reality.

A recent study [75] is to detect and analyze moving dynamic spatio-temporal regions and their mobility in large sensor datasets. This kind of region often implies locally intense areas of precipitation, anomalous sea surface temperature readings, and locally high levels of water pollution, etc. It can also be regarded as mining group patterns of a phenomenon.

E. URBAN SERVICE

Knowledge discovered with trajectory data mining techniques helps to improve quality of life in urban areas from several aspects [50], [51], [76]–[79].

Through analyzing a large scale of trajectory data collected from electronic vehicles, Li *et al.* [76] solve a challenging question of how to strategically deploy charging stations and charging points, thus minimizing average time to the nearest charging station and average waiting time for an available charging point.

Inferring road maps from large-scale GPS traces are highly promising and attractive, since building maps from geographical surveys are expensive and infrequent. Liu *et al.* [77] address a problem of map inference in a practical setting, i.e., GPS traces has very low resolution and sampling frequency. Several techniques for map inference from sparse data are investigated and extensively evaluated.

Traffic volume estimation is a primary task in many applications, such as risk analysis, quality of service, location ranking. A recent study [78] aims to estimate traffic volume for pedestrians within closed environments. Knowledge on people's presence provides a valuable opportunity for improving infrastructure, e.g., locations of information desks, shops or toilets, path-widths of corridors in a stadium.

Parking service is of great importance to citizens in urban areas. Parking places (especially on-street parking) are usually unavailable in existing electronic maps. iPark [79] aims to enable parking search applications and to provide complete parking information, i.e., annotating an existing map with parking zones based on trajectory data of vehicles.

A developed city naturally has different functional regions, e.g., residential areas, business districts, and educational areas. The knowledge is highly valuable to both citizens and urban planners. People living a city need the knowledge to assist their decision on buying or renting a house, choosing a job. Meanwhile, the knowledge helps urban planners to make decisions on future development of the city and to estimate effects of previous policies. Yuan *et al.* [50], [51] address a problem of discovering regions of different functions in a city based on a large scale of trajectory data. A topic model based approach has been proposed to cluster segmented regions into functional zones, where a region is regarded as a document and a function as a topic.

F. MAKING SENSE OF TRAJECTORIES

Raw trajectory data which are in the form of sequence of geographical locations and timestamps fail to make sense to people without semantic description. There exist a great many studies [7], [80]–[83] to facilitate interpretation of raw trajectory data.

Unlike semantic trajectory that cannot express movement properties of moving objects, e.g., overspeed, stopover, Su *et al.* [7] propose a partition-and-summarization approach that automatically generates a short human-readable text to describe a trajectory. The approach not only extends expressivity of traditional semantic trajectories but also avoids a challenging problem of storage, processing and transmission of large volume of semantic trajectories. A raw trajectory data is first segmented according to behavior of a moving object, and then characteristics of each trajectory segmentation are summarized by short textual description. Furthermore, a proto system named STMaker [80] based on this idea is implemented.

It is certainly worth noting that semantic meaning of locations and short textual messages collected by social media services provide an unprecedented opportunity to interpret raw trajectory data. TOPTRAC [81] aims to detect latent topic in trajectory data. Specifically, the approach not only finds semantic regions with a coherent topic but also extracts mobility patterns of human beings between semantic regions. Similarly, Lu *et al.* [83] employ a clustering-based approach to discover semantic regions.

A lot of emerging location-aware applications require a semantic notation of a location point, e.g., "home", "work", instead of latitude and longitude coordinates. Lv *et al.* [82] propose a method of automatically discovering personal semantic places (i.e., both a physical location and semantic meaning of the location).

X. OPEN ISSUES

In spite of its various applications, trajectory data mining techniques must be improved from many aspects. We offer a few open issues in the following. First, although current trajectory data mining techniques help to analyze behavior of moving objects, we have limited understanding of root causes of such interesting behavior at all. Second, current privacy-preserving methods are far from enough. Privacy-preserving is of great importance to trajectory data sharing and publication. Third, it is possible to extract much more value if trajectory data are combined with other sources of data, e.g., healthcare data. For instance, analyzing correlation between one's historical trajectories and his or her illness may provide clues for causes of the illness.

XI. CONCLUSION

Trajectory data mining is beneficial to individual citizens. One can understand his or her movement behavior better through analyzing historical trajectories. Besides, trajectory data mining provides plenty of convenience to the public, e.g., route recommendation, real-time traffic information publication by transport agencies. However, people suffer from privacy breaches if their trajectories are collected and utilized inappropriately. Moreover, people are usually disturbed by commercial advertisements which are possibly pushed in the name of personalized services.

For the government and some organizations, trajectory data mining helps to reduce cost of supervision and management. In urban areas, trajectory data mining from vehicle trajectories provides an efficient and scalable method to monitor traffic condition of the whole city. Another example is to record illegal or irregular behavior which is probably valuable to ascertain responsibilities later. For example, overspeed can be inferred from trajectories. This evidence is valuable especially in roads without roadside cameras. Similarly, commercial organizations expect to cut down their costs in virtue of trajectory data mining. For example, RFID data, as a special kind of trajectories, indeed help to manage commodity stocks.

Location acquisition technologies generate huge amount of trajectory data. Trajectory data which track traces of moving objects is typically represented by a sequence of timestamped geographical locations. A large amount of applications are created upon mining trajectory data. The survey reviews an extensive collection of existing studies in the proposed framework of trajectory data mining.

REFERENCES

[1] H. Jeung, H. Lu, S. Sathe, and M. L. Yiu, "Managing evolving uncertainty in trajectory databases," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 7, pp. 1692–1705, Jul. 2014.

[2] R. G. Baraniuk, "More is less: Signal processing and the data deluge," *Science*, vol. 331, no. 6018, pp. 717–719, Feb. 2011.

[3] S. Liu, J. Pu, Q. Luo, H. Qu, L. M. Ni, and R. Krishnan, "VAIT: A visual analytics system for metropolitan transportation," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 4, pp. 1586–1596, Dec. 2013.

[4] Y. Ge, H. Xiong, Z.-H. Zhou, H. Ozdemir, J. Yu, and K. C. Lee, "Top-Eye: Top-k evolving trajectory outlier detection," in *Proc. 19th ACM Conf. Inf. Knowl. Manage. (CIKM)*, Toronto, ON, Canada, Oct. 2010, pp. 1733–1736.

[5] B. Fazzinga, S. Flesca, F. Furfaro, and F. Parisi, "Cleaning trajectory data of RFID-monitored objects through conditioning under integrity constraints," in *Proc. 17th Int. Conf. Extending Database Technol. (EDBT)*, Athens, Greece, Mar. 2014, pp. 379–390.

[6] N. Pelekis, I. Kopanakis, C. Panagiotakis, and Y. Theodoridis, "Unsupervised trajectory sampling," in *Proc. Mach. Learn. Knowl. Discovery Databases, Eur. Conf. (ECML PKDD)*, Barcelona, Spain, Sep. 2010, pp. 17–33.

[7] H. Su *et al.*, "Making sense of trajectory data: A partition-and-summarization approach," in *Proc. 31st IEEE Int. Conf. Data Eng. (ICDE)*, Seoul, South Korea, Apr. 2015, pp. 963–974.

[8] H. Wang, K. Zheng, J. Xu, B. Zheng, X. Zhou, and S. W. Sadiq, "SharkDB: An in-memory column-oriented trajectory storage," in *Proc. 23rd ACM Int. Conf. Inf. Knowl. Manage. (CIKM)*, Shanghai, China, Nov. 2014, pp. 1409–1418.

[9] H. Wang, K. Zheng, X. Zhou, and S. Sadiq, "SharkDB: An in-memory storage system for massive trajectory data," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, Melbourne, VIC, Australia, May/Jun. 2015, pp. 1099–1104.

[10] P. Banerjee, S. Ranu, and S. Raghavan, "Inferring uncertain trajectories from partial observations," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Shenzhen, China, Dec. 2014, pp. 30–39.

[11] M. Li, A. Ahmed, and A. J. Smola, "Inferring movement trajectories from GPS snippets," in *Proc. 8th ACM Int. Conf. Web Search Data Mining (WSDM)*, Shanghai, China, Feb. 2015, pp. 325–334.

[12] M. Chiang, Y.-H. Lin, W.-C. Peng, and P. S. Yu, "Inferring distant-time location in low-sampling-rate trajectories," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, Chicago, IL, USA, Aug. 2013, pp. 1454–1457.

[13] K. Zheng, Y. Zheng, X. Xie, and X. Zhou, "Reducing uncertainty of low-sampling-rate trajectories," in *Proc. IEEE 28th Int. Conf. Data Eng. (ICDE)*, Washington, DC, USA, Apr. 2012, pp. 1144–1155.

[14] N. D. Larusso and A. Singh, "Efficient tracking and querying for coordinated uncertain mobile objects," in *Proc. 29th IEEE Int. Conf. Data Eng. (ICDE)*, Brisbane, QLD, Australia, Apr. 2013, pp. 182–193.

[15] H. Su, K. Zheng, H. Wang, J. Huang, and X. Zhou, "Calibrating trajectory data for similarity-based analysis," in *Proc. ACM SIGMOD Int. Conf. Manage. Data (SIGMOD)*, New York, NY, USA, Jun. 2013, pp. 833–844.

[16] H. Su, K. Zheng, J. Huang, H. Wang, and X. Zhou, "Calibrating trajectory data for spatio-temporal similarity analysis," *VLDB J.*, vol. 24, no. 1, pp. 93–116, Feb. 2015.

[17] C. Panagiotakis, N. Pelekis, I. Kopanakis, E. Ramasso, and Y. Theodoridis, "Segmentation and sampling of moving object trajectories based on representativeness," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 7, pp. 1328–1343, Jul. 2012.

[18] Y. Li *et al.*, "Sampling big trajectory data," in *Proc. 24th ACM Int. Conf. Inf. Knowl. Manage. (CIKM)*, Melbourne, VIC, Australia, Oct. 2015, pp. 941–950.

[19] A. Nibali and Z. He, "Trajic: An effective compression system for trajectory data," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 11, pp. 3138–3151, Nov. 2015.

[20] R. Song, W. Sun, B. Zheng, and Y. Zheng, "PRESS: A novel framework of trajectory compression in road networks," in *Proc. PVLDB*, 2014, vol. 7. no. 9, pp. 661–672.

[21] J. Liu, K. Zhao, P. Sommer, S. Shang, B. Kusy, and R. Jurdak, "Bounded quadrant system: Error-bounded trajectory compression on the go," in *Proc. 31st IEEE Int. Conf. Data Eng. (ICDE)*, Seoul, South Korea, Apr. 2015, pp. 987–998.

[22] C. Long, R. C.-W. Wong, and H. V. Jagadish, "Direction-preserving trajectory simplification," in *Proc. PVLDB*, Aug. 2013, vol. 6, no. 10, pp. 949–960.

[23] C. Long, R. C.-W. Wong, and H. V. Jagadish, "Trajectory simplification: On minimizing the direction-based error," in *Proc. PVLDB*, 2014, vol. 8. no. 1, pp. 49–60.

[24] P. Cudré-Mauroux, E. Wu, and S. Madden, "Trajstore: An adaptive storage system for very large trajectory data sets," in *Proc. 26th Int. Conf. Data Eng. (ICDE)*, Mar. 2010, Long Beach, CA, USA, 2010, pp. 109–120.

[25] S. Ranu, D. P, A. D. Telang, P. Deshpande, and S. Raghavan, "Indexing and matching trajectories under inconsistent sampling rates," in *Proc. 31st IEEE Int. Conf. Data Eng. (ICDE)*, Seoul, South Korea, Apr. 2015, pp. 999–1010.

[26] I. S. Popa, K. Zeitouni, V. Oria, D. Barth, and S. Vial, "Indexing in-network trajectory flows," *VLDB J.*, vol. 20, no. 5, pp. 643–669, 2011.

[27] J. Ni and C. V. Ravishankar, "Indexing spatio-temporal trajectories with efficient polynomial approximations," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 5, pp. 663–678, May 2007.

[28] Z. Chen, H. T. Shen, X. Zhou, Y. Zheng, and X. Xie, "Searching trajectories by locations: An efficiency study," in *Proc. ACM SIGMOD Int. Conf. Manage. Data (SIGMOD)*, Indianapolis, IN, USA, Jun. 2010, pp. 255–266.

[29] D. Yan, J. Cheng, Z. Zhao, and W. Ng, "Efficient location-based search of trajectories with location importance," *Knowl. Inf. Syst.*, vol. 45, no. 1, pp. 215–245, Oct. 2015.

[30] K. Zheng, G. Trajcevski, X. Zhou, and P. Scheuermann, "Probabilistic range queries for uncertain trajectories on road networks," in *Proc. 14th Int. Conf. Extending Database Technol. (EDBT)*, Uppsala, Sweden, Mar. 2011, pp. 283–294.

[31] L. Zhan, Y. Zhang, W. Zhang, X. Wang, and X. Lin, "Range search on uncertain trajectories," in *Proc. 24th ACM Int. Conf. Inf. Knowl. Manage. (CIKM)*, Melbourne, VIC, Australia, Oct. 2015, pp. 921–930.

[32] R. H. Güting, T. Behr, and J. Xu, "Efficient k-nearest neighbor search on moving object trajectories," *VLDB J.*, vol. 19, no. 5, pp. 687–714, Oct. 2010.

[33] J. Niedermayer, A. Züfle, T. Emrich, M. Renz, N. Mamoulis, L. Chen, and H.-P. Kriegel, "Probabilistic nearest neighbor queries on uncertain moving object trajectories," in *Proc. PVLDB*, 2013, vol. 7, no. 3, pp. 205–216.

[34] C. Ma, H. Lu, L. Shou, and G. Chen, "KSQ: Top-k similarity query on uncertain trajectories," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 9, pp. 2049–2062, Sep. 2013.

[35] M. R. Vieira, P. Bakalov, and V. J. Tsotras, "Querying trajectories using flexible patterns," in *Proc. 13th Int. Conf. Extending Database Technol. (EDBT)*, Lausanne, Switzerland, Mar. 2010, pp. 406–417.

[36] B. Zheng, N. J. Yuan, K. Zheng, X. Xie, S. W. Sadiq, and X. Zhou, "Approximate keyword search in semantic trajectory database," in *Proc. 31st IEEE Int. Conf. Data Eng. (ICDE*, Seoul, South Korea, Apr. 2015, pp. 975–986.

[37] K. Zheng, S. Shang, N. J. Yuan, and Y. Yang, "Towards efficient search for activity trajectories," in *Proc. 29th IEEE Int. Conf. Data Eng. (ICDE)*, Brisbane, QLD, Australia, Apr. 2013, pp. 230–241.

[38] S. Shang, R. Ding, B. Yuan, K. Xie, K. Zheng, and P. Kalnis, "User oriented trajectory search for trip recommendation," in *Proc. 15th Int. Conf. Extending Database Technol. (EDBT)*, Berlin, Germany, Mar. 2012, pp. 156–167.

[39] S. Shang, R. Ding, K. Zheng, C. S. Jensen, P. Kalnis, and X. Zhou, "Personalized trajectory matching in spatial networks," *VLDB J.*, vol. 23, no. 3, pp. 449–468, Jun. 2014.

[40] K. Zheng, Y. Zheng, N. J. Yuan, and S. Shang, "On discovery of gathering patterns from trajectories," in *Proc. 29th IEEE Int. Conf. Data Eng. (ICDE)*, Brisbane, QLD, Australia, Apr. 2013, pp. 242–253.

[41] K. Zheng, Y. Zheng, N. J. Yuan, S. Shang, and X. Zhou, "Online discovery of gathering patterns over trajectories," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 8, pp. 1974–1988, Aug. 2014.

[42] X. Li, V. Ceikute, C. S. Jensen, and K. Tan, "Effective online group discovery in trajectory databases," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 12, pp. 2752–2766, Dec. 2013.

[43] C. Zhang, J. Han, L. Shou, J. Lu, and T. La Porta, "Splitter: Mining fine-grained sequential patterns in semantic trajectories," in *Proc. PVLDB*, 2014, vol. 7. no. 9, pp. 769–780.

[44] Z. Li, B. Ding, J. Han, R. Kays, and P. Nye, "Mining periodic behaviors for moving objects," in *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Washington, DC, USA, Jul. 2010, pp. 1099–1108. [Online]. Available: http://doi.acm.org/10.1145/1835804.1835942

[45] Z. Li, J. Han, B. Ding, and R. Kays, "Mining periodic behaviors of object movements for animal and biological sustainability studies," *Data Mining Knowl. Discovery*, vol. 24, no. 2, pp. 355–386, Mar. 2012. [Online]. Available: http://dx.doi.org/10.1007/s10618-011-0227-9

[46] G. de Vries and M. van Someren, "Clustering vessel trajectories with alignment kernels under trajectory compression," in *Proc. Eur. Conf. Mach. Learn. Knowl. Discovery Databases (ECML PKDD)*, Barcelona, Spain, Sep. 2010, pp. 296–311.

[47] S. Liu, S. Wang, K. Jayarajah, A. Misra, and R. Krishnan, "TODMIS: Mining communities from trajectories," in *Proc. 22nd ACM Int. Conf. Inf. Knowl. Manage. (CIKM)*, San Francisco, CA, USA, Oct./Nov. 2013, pp. 2109–2118.

[48] S. Liu, Y. Liu, L. M. Ni, J. Fan, and M. Li, "Towards mobility-based clustering," in *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Washington, DC, USA, Jul. 2010, pp. 919–928.

[49] D. Patel, C. Sheng, W. Hsu, and M. L. Lee, "Incorporating duration information for trajectory classification," in *Proc. IEEE 28th Int. Conf. Data Eng. (ICDE)*, Washington, DC, USA, Apr. 2012, pp. 1132–1143.

[50] N. J. Yuan, Y. Zheng, X. Xie, Y. Wang, K. Zheng, and H. Xiong, "Discovering urban functional zones using latent activity trajectories," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 3, pp. 712–725, Mar. 2015.

[51] J. Yuan, Y. Zheng, and X. Xie, "Discovering regions of different functions in a city using human mobility and POIs," in *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, Beijing, China, Aug. 2012, pp. 186–194.

[52] X. Wang, G. Li, G. Jiang, and Z. Shi, "Semantic trajectory-based event detection and event pattern mining," *Knowl. Inf. Syst.*, vol. 37, no. 2, pp. 305–329, Nov. 2013.

[53] T. S. Au, R. Duan, H. Kim, and G. Q. Ma, "Spatiotemporal event detection in mobility network," in *Proc. 10th IEEE Int. Conf. Data Mining (ICDM)*, Sydney, NSW, Australia, Dec. 2010, pp. 28–37.

[54] N. Andrienko, G. Andrienko, G. Fuchs, and P. Jankowski, "Visual analytics methodology for scalable and privacy-respectful discovery of place semantics from episodic mobility data," in *Proc. Eur. Conf. Mach. Learn. Knowl. Discovery Databases (ECML PKDD)*, Porto, Portugal, Sep. 2015, pp. 254–258.

[55] C. Costa, C. Laoudias, D. Zeinalipour-Yazti, and D. Gunopulos, "SmartTrace: Finding similar trajectories in smartphone networks without disclosing the traces," in *Proc. 27th Int. Conf. Data Eng. (ICDE)*, Hanover, Germany, Apr. 2011, pp. 1288–1291.

[56] N. Pelekis, A. Gkoulalas-Divanis, M. Vodas, D. Kopanaki, and Y. Theodoridis, "Privacy-aware querying over sensitive trajectory data," in *Proc. 20th ACM Conf. Inf. Knowl. Manage. (CIKM)*, Glasgow, Scotland, Oct. 2011, pp. 895–904.

[57] L. Kong, L. He, X. Y. Liu, Y. Gu, M. Wu, and X. Liu, "Privacy-preserving compressive sensing for crowdsensing based trajectory recovery," in *Proc. 35th IEEE Int. Conf. Distrib. Comput. Syst. (ICDCS)*, Columbus, OH, USA, Jun./Jul. 2015, pp. 31–40. [Online]. Available: http://dx.doi.org/10.1109/ICDCS.2015.12

[58] E. H.-C. Lu, W.-C. Lee, and V. S. Tseng, "Mining fastest path from trajectories with multiple destinations in road networks," *Knowl. Inf. Syst.*, vol. 29, no. 1, pp. 25–53, Oct. 2011.

[59] Z. Chen, H. T. Shen, and X. Zhou, "Discovering popular routes from trajectories," in *Proc. 27th Int. Conf. Data Eng. (ICDE)*, Hanover, Germany, Apr. 2011, pp. 900–911.

[60] L.-Y. Wei, Y. Zheng, and W.-C. Peng, "Constructing popular routes from uncertain trajectories," in *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, Beijing, China, Aug. 2012, pp. 195–203.

[61] W. Luo, H. Tan, L. Chen, and L. M. Ni, "Finding time period-based most frequent path in big trajectory data," in *Proc. ACM SIGMOD Int. Conf. Manage. Data (SIGMOD)*, New York, NY, USA, Jun. 2013, pp. 713–724.

[62] Y. Liu *et al.*, "Exploiting heterogeneous human mobility patterns for intelligent bus routing," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Shenzhen, China, Dec. 2014, pp. 360–369.

[63] J. Dai, B. Yang, C. Guo, and Z. Ding, "Personalized route recommendation using big trajectory data," in *Proc. 31st IEEE Int. Conf. Data Eng. (ICDE)*, Seoul, South Korea, Apr. 2015, pp. 543–554.

[64] A. Y. Xue, R. Zhang, Y. Zheng, X. Xie, J. Huang, and Z. Xu, "Destination prediction by sub-trajectory synthesis and privacy protection against such prediction," in *Proc. 29th IEEE Int. Conf. Data Eng. (ICDE)*, Brisbane, QLD, Australia, Apr. 2013, pp. 254–265.

[65] A. Y. Xue, R. Zhang, Y. Zheng, X. Xie, J. Yu, and Y. Tang, "DesTeller: A system for destination prediction based on trajectories with privacy protection," in *Proc. PVLDB*, 2013, vol. 6. no. 12, pp. 1198–1201.

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

[67] Y. Wang *et al.*, "Regularity and conformity: Location prediction using heterogeneous mobility data," in *Proc. 21st ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Sydney, NSW, Australia, Aug. 2015, pp. 1275–1284.

[68] C. Renso, M. Baglioni, J. A. F. de Macêdo, R. Trasarti, and M. Wachowicz, "How you move reveals who you are: Understanding human behavior by analyzing trajectory data," *Knowl. Inf. Syst.*, vol. 37, no. 2, pp. 331–362, Nov. 2013.

[69] X. Song, Q. Zhang, Y. Sekimoto, and R. Shibasaki, "Prediction of human emergency behavior and their mobility following large-scale disaster," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, New York, NY, USA, Aug. 2014, pp. 5–14.

[70] S. Ando and E. Suzuki, "Role-behavior analysis from trajectory data by cross-domain learning," in *Proc. 11th IEEE Int. Conf. Data Mining (ICDM)*, Vancouver, BC, Canada, Dec. 2011, pp. 21–30.

[71] H. Gao, J. Tang, X. Hu, and H. Liu, "Modeling temporal effects of human mobile behavior on location-based social networks," in *Proc. 22nd ACM Int. Conf. Inf. Knowl. Manage. (CIKM)*, San Francisco, CA, USA, Oct./Nov. 2013, pp. 1673–1678.

[72] S. Liu, Q. Qu, and S. Wang, "Rationality analytics from trajectories," *ACM Trans. Knowl. Discovery Data*, vol. 10, no. 1, p. 10, Jul. 2015.

[73] M. C. González, C. A. Hidalgo, and A.-L. Barabási, "Understanding individual human mobility patterns," *Nature*, vol. 453, pp. 779–782, Jun. 2008. [Online]. Available: http://dx.doi.org/10.1038/nature06958

[74] A. Gupta *et al.*, "A mobility simulation framework of humans with group behavior modeling," in *Proc. IEEE 13th Int. Conf. Data Mining*, Dallas, TX, USA, Dec. 2013, pp. 1067–1072.

[75] M. P. McGuire, V. P. Janeja, and A. Gangopadhyay, "Mining trajectories of moving dynamic spatio-temporal regions in sensor datasets," *Data Mining Knowl. Discovery*, vol. 28, no. 4, pp. 961–1003, Jul. 2014.

[76] Y. Li, J. Luo, C. Y. Chow, K. L. Chan, Y. Ding, and F. Zhang, "Growing the charging station network for electric vehicles with trajectory data analytics," in *Proc. 31st IEEE Int. Conf. Data Eng. (ICDE)*, Seoul, South Korea, Apr. 2015, pp. 1376–1387.

[77] X. Liu, J. Biagioni, J. Eriksson, Y. Wang, G. Forman, and Y. Zhu, "Mining large-scale, sparse GPS traces for map inference: Comparison of approaches," in *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, Beijing, China, Aug. 2012, pp. 669–677.

[78] T. Liebig, Z. Xu, M. May, and S. Wrobel, "Pedestrian quantity estimation with trajectory patterns," in *Proc. Eur. Conf. Mach. Learn. Knowl. Discovery Databases (ECML PKDD)*, Bristol, U.K., Sep. 2012. pp. 629–643.

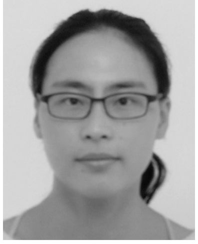
[79] B. Yang, N. Fantini, and C. S. Jensen, "iPark: Identifying parking spaces from trajectories," in *Proc. 16th Int. Conf. Extending Database Technol. (EDBT)*, Genoa, Italy, Mar. 2013, pp. 705–708.

[80] H. Su, K. Zheng, K. Zheng, J. Huang, and X. Zhou, "STMaker: A system to make sense of trajectory data," in *Proc. PVLDB*, 2014, vol. 7. no. 13, pp. 1701–1704.

[81] Y. Kim, J. Han, and C. Yuan, "TOPTRAC: Topical trajectory pattern mining," in *Proc. 21st ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Sydney, NSW, Australia, Aug. 2015, pp. 587–596.

[82] M. Lv, L. Chen, and G. Chen, "Discovering personally semantic places from GPS trajectories," in *Proc. 21st ACM Int. Conf. Inf. Knowl. Manage. (CIKM)*, Maui, HI, USA, Oct./Nov. 2012, pp. 1552–1556.

[83] C. Lu, P.-R. Lei, W.-C. Peng, and I.-J. Su, "A framework of mining semantic regions from trajectories," in *Proc. 16th Int. Conf. Database Syst. Adv. Appl. (DASFAA)*, Hong Kong, Apr. 2011, pp. 193–207.



ZHENNI FENG received the B.Eng. degree in computer science from the Huazhong University of Science and Technology, Wuhan, China, in 2012. She is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering, Shanghai Jiao Tong University. Her research interests include machine learning, spatiotemporal data mining, and big data analysis.

YANMIN ZHU (M'05) received the B.Eng. degree in computer science from Xi'an Jiaotong University, in 2002, and the Ph.D. degree in computer science from the Hong Kong University of Science and Technology, in 2007. He was a Research Associate with the Department of Computing, Imperial College London. He is currently a Professor with the Department of Computer Science and Engineering, Shanghai Jiao Tong University. His research interests include wireless sensor network, mobile computing, and big data analysis and systems. He is a member of the IEEE Communication Society.