IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received September 3, 2018, accepted September 23, 2018, date of publication October 4, 2018, date of current version October 31, 2018.

*Digital Object Identifier 10.1109/ACCESS.2018.2873779*

# Big Trajectory Data: A Survey of Applications and Services

**XIANGJIE KONG, (Senior Member, IEEE), MENGLIN LI, KAI MA, KAIQI TIAN, MENGYUAN WANG, ZHAOLONG NING, (Member, IEEE), AND FENG XIA, (Senior Member, IEEE)**

Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province, School of Software, Dalian University of Technology, Dalian 116620, China

Corresponding author: Feng Xia (f.xia@ieee.org)

This work was supported in part by the National Natural Science Foundation of China under Grant 61572106 and Grant 61502075, in part by the Dalian Science and Technology Innovation Fund under Grant 2018J12GX048, and in part by the Fundamental Research Funds for the Central Universities under Grant DUT18JC09 and Grant DUT17RC(4)49.

**ABSTRACT** The rapid development of wireless infrastructure and data acquisition technologies contributes to the explosive growth of data, especially trajectory data with rich information. Trajectory data, which records locations of moving objects at certain moments, has long been an important means of studying human behavior and solving traffic problems. In this paper, we mainly introduce the trajectory data from the perspective of applications and services. According to the degree of data structured, we divide the trajectory data into explicit trajectory data and implicit trajectory data, and describe each type in detail. Then, we introduce the applications of trajectory data from travel behavior, travel patterns, and other aspects. Combined with case studies, we provide a description to the services of trajectory data in terms of transportation administration and commercial service. Finally, we focus on challenges in trajectory data, including privacy protection, human mobility causality, and emission reduction.

**INDEX TERMS** Trajectory data, human mobility, travel behavior, applications and services.

## I. INTRODUCTION

As the information and communication technology develops rapidly, tremendous volumes of data which are captured by various Internet of Things (IoT) devices have experienced exponential growth. In IoT, every data acquisition device is placed at a specific geographic location and every piece of data has a time stamp. Time and space are important dimensions for statistical analysis and their correlation is an important property of data from IoT [1]. Apart from IoT services, like Global Position Systems (GPS), Radio Frequency Identification (RFID), Automated Fare Collection Systems (AFC), GSM beacons, social network data also contains time and location information to generate trajectory data.

Trajectory data analysis is of significant practical value. There are numerous applications and services of trajectory data for the government, commercial organizations, and individuals. For the government, trajectory data analysis helps it to reduce the costs of management and to establish reasonable strategies for urban planning, e.g., monitoring irregular behaviors of vehicles, such as overspeed and reverse driving. Even crime behaviors can be inferred from trajectories. In urban services, bus and taxi services are vital for citizens' commuting, and those public services can be optimized by analyzing historical trajectories. For commercial organizations, like Didi and Uber, trajectory data analysis can help understand users' behaviors to satisfy their needs and to enhance commercial competitiveness. Moreover, personalized services based on trajectories are of great concern for customer satisfaction. For individuals, we can understand residents' behaviors well by analyzing historical trajectory data. Some trajectory-based services provided by commercial organizations bring plenty of conveniences to our daily life. For example, the real-time traffic detection service ensures that we can save our time with an optimal route when a traffic jam occurs. However, people suffer from privacy problems, like annoying advertisements from unknown sources, if trajectory data is collected inappropriately.

This paper aims to summarize the applications and services of trajectory data. The rest of the paper is structured as follows. Section II offers the concepts of trajectory data and its classifications. A detailed review of trajectory applications is provided in Section III. Then the services of trajectory data are described in Section IV. In Section V, the challenges of

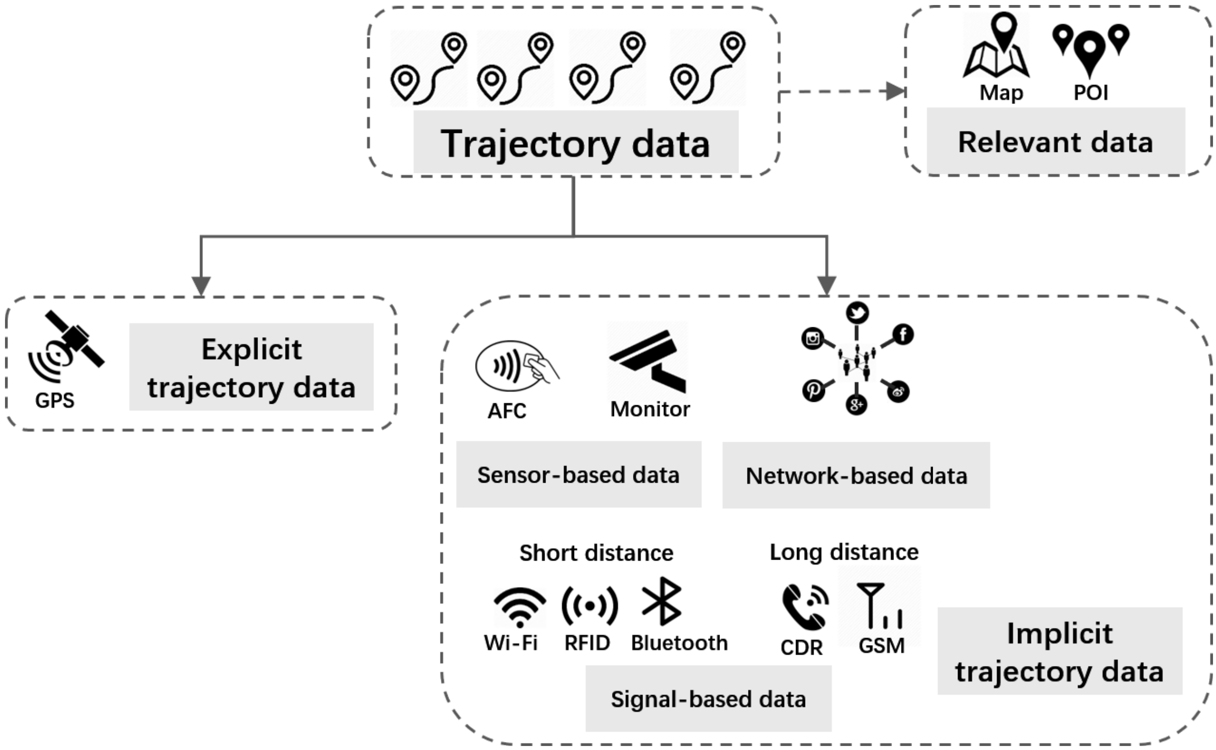


**FIGURE 1.** Classifications of trajectory data.

trajectory data are discussed. Finally, we conclude this paper in Section VI.

## II. TRAJECTORY DATA

What is trajectory data? As Zheng [2] defines, the spatial trajectory is a trace generated by a moving object in geographical spaces and is usually represented by a series of chronologically ordered points. Mazimpaka and Timpf also provide a definition of trajectory data that a trajectory of a moving object is a discrete trace that the moving object travels in geographical spaces [3].

In a variety of data, certain types of data display trajectories of persons, animals, objects and so forth, like GPS data. There exists such data containing rich spatial and temporal information. Although they are not the trajectory data we usually think of, we can extract trajectories from them by basic data processing operations, such as social data with geotags. Therefore, trajectory data can be roughly categorized into explicit trajectory data and implicit trajectory data, as shown in Fig.1. As the representative of explicit trajectory data, GPS data records positions of objects continuously and intensively at uniform time intervals. However, for implicit trajectory data, the time granularity is relatively large and the distribution of recorded time points is relatively random, which is closely related to data sources. In this section, we introduce the two categories of trajectory data and several types of data that are quite relevant to trajectory data.

### A. EXPLICIT TRAJECTORY DATA

We define explicit trajectory data as a type of well-structured data which directly provides time and location information and have quite strong spatiotemporal continuity. Explicit trajectory data is based on time and locations and GPS data is the most representative one, which is widely used in researches of trajectory analysis. Each record of GPS data contains time, latitude and longitude information, and other additional information, like speed [4].

### B. IMPLICIT TRAJECTORY DATA

Different from the explicit trajectory data, implicit trajectory data has the weak spatiotemporal continuity. Moreover, explicit trajectory data usually has multiple storage formats. We classify explicit trajectory data into signal-based, sensor-based, and network-based on the fundamental of data sources.

#### 1) SIGNAL-BASED DATA

Signal-based data collection requires multiple signal sources to be distributed in different locations, e.g., routers for wifi, base stations for GSM, CDR, and IoT equipment for Bluetooth and RFID. Receiving devices like mobile phones are also required. A single signal source corresponds to multiple receiving devices. What's more, for signal-based data, the range of data acquisition is wide, while the location accuracy is relatively poor. The fields of signal-based data tend to be complex, such as device identification, connection/disconnection time, signal strength and additional information. Based on signal transmission distance, signal-based data can be further divided into short distances, including wifi data, RFID, and Bluetooth data, and long distance, containing CDR and GSM.

**TABLE 1.** The summary of applications of trajectory data in travel behavior, travel pattern, urban planning, and others.

| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| **Travel Behaviors** | | | | | | | |
| Human Mobility Pattern | [6], [7], [8], [9], [10], [11], [12] | [13], [14], [15] | [16], [8] | [17], [16], [9], [18] | [6] | [8] | [17], [11] |
| Activities Prediction | [19], [20], [21] | [19] | | [22], [23], [24] | [25] | [22], [20], [21] | [19], [20] |
| Anomalous Detection | [26], [27] | [28], [26], [29], [30] | | [31] | | | [27] |
| **Travel Pattern** | | | | | | | |
| Trip Purpose Estimation | [32], [33] | [34] | | [35], [36] | | [37], [32], [33], [36] | [34] |
| Destination Prediction | [38], [39], [40], [41] | | | [39], [41] | | | [39] |
| Route Discovery | [42], [43], [44], [45] | [46] | | [42], [43] | | [43] | [43] |
| Travel Modes Analysis | [47], [48], [49], [50] | | | | | [48] | [47], [49] |
| **Others** | | | | | | | |
| Urban Functions | [51], [52] | [53], [51], [54] | | | [55] | [53], [55], [52] | |
| Time Inference | [56], [57], [58], [59] | | [60] | | | [57] | [57], [59] |
| Environmental Monitoring | [61], [62], [63], [64], [65] | | | [63], [66] | | | [67], [68], [65] |

**A**: Explicit trajectory data (GPS) **B**: Sensor-based data **C**: Signal-based data (short distance) **D**: Signal-based data (long distance) **E**: Network-based data **F**: POI **G**: Map

#### 2) SENSOR-BASED DATA

Sensor-based data is recorded when the object is passing by the sensor. For example, the traffic monitor can collect location information of passing vehicles. Due to the physical location limit of sensors, the scope of data acquisition is small and the location accuracy is relatively high. Sensor-based data records object identity, passing-by time and other additional information. As a representative sensor-based data, Automatic Fare Collection (AFC) collect smart card transaction data and is used to analyze or to improve public transportation [5].

#### 3) NETWORK-BASED DATA

With the rise of social sites in recent years, growing social data with geotags can be obtained, which provides a new idea for analyzing trajectory data. Such kind of data is uploaded by users and obtained from the entire internet. Thus, we define it as network-based data. The data from Facebook and Weibo is the representative of network-based data. Network-based data is interesting and can offer additional semantic information of events, however, due to the high dependence on users' behaviors, there are lots of noises and complex data preprocessing.

### C. RELEVANT DATA

Trajectory data contains a wealth of information. However, in the analysis of trajectory data, researchers usually integrate it with other data to achieve effective mining of trajectory data. Here we introduce two common relevant data, POI (Point of Interests) data and map data.

POI contains information of actual buildings, such as hotels, hospitals, supermarkets, and stations. With the application of POI data, we can transform the longitude and latitude information into meaningful building information and provide a reasonable explanation of patterns and phenomena discovered in trajectory data. The region of interest (ROI), the area of interest (AOI) and the volume of interest (VOI) are highly relevant to POI data.

The trajectory data contains a large amount of geographic information. Urban road networks and terrain distribution have a huge impact on the formation of trajectories of moving objects such as persons and vehicles. Therefore, it is a natural idea and practice to incorporate the trajectory data into a map, which contains rich information like road networks to view the distribution of data. The importance of map data for trajectory data analysis is not negligible. Google Map, Leaflet API and Mapbox API are typically practical map tools.

## III. APPLICATIONS OF TRAJECTORY DATA

The trajectory data has high application value in many fields. In this section, we provide a comprehensive demonstration of the application of trajectory data from three main aspects, human behavior, travel pattern, and others.

As shown in Table 1, we summarize the applications of trajectory data in human behavior, travel pattern, urban planning, and others. From the perspective of the classifications of trajectory data, we can find that the explicit trajectory data, GPS, is widely used in the applications of trajectory data because of its clear structure and rich spatiotemporal information. Thanks to the characteristic of easy collections, the utilization of sensor-based data and signal-based data (long distance) are also relatively high in applications. However, for signal-based data (short distance), such as wifi data and Bluetooth data, the characteristic of small scope limits its range of applications to just human mobility pattern analysis and time inference. Compared with other applications, anomaly detection uses sensor-based data, especially video monitoring data, having high application value. Activities prediction, trip purpose estimation, and urban function, highly depend on POI data. And almost all trajectory application count on map data.

### A. TRAVEL BEHAVIORS

As one of the most important research topics, human travel behavior has attracted the attention of many studies because of its complexity and variability for a long time. The rich information hidden in trajectory data can help understand human travel behavior and mine interesting mobility patterns to improve the quality of urban lives. We introduce human behavior analysis of trajectory data from three aspects: human mobility pattern, activity prediction, and anomalous event detection.

#### 1) HUMAN MOBILITY PATTERN

Human mobility pattern analysis is an important sub-area of human behavior analysis and it can yield insight into multiple issues like urban planning. Links among individuals are highly predictable, but the factors which mostly affect the human mobility are ambiguous. Giannotti *et al.* [6] provide a complexity science perspective on human mobility to analyze multiple variables of human mobility pattern and to predict the possibility of people's movements. Shaw *et al.* [9] introduce human dynamics and the relations with trajectories in the mobile and big data era in detail.

Based on specific influencing factors of human mobility and the clear target of analysis, researchers propose effective methodologies and frameworks to understand human mobility pattern. Renso *et al.* [7] present a semantic-enriched knowledge discovery process to estimate the human mobility and make it meaningful, which is evaluated in the application domain of traffic management and recreation behavior. Qiao *et al.* [18] propose a mobility analytical framework to analyze massive data traffic from the mobile Internet, and validate the framework with a few common rules.

Human mobility is a broad concept and contains short-term, median-term and long-term travel behaviors. The short-term and median-term travel behaviors have high contingency and variability, which makes them more difficult to study than the long-term travel behaviors. An example of long-term study is worker commuting. Zhou *et al.* [13] conduct a detailed research about how to increase the efficiency of commuting using smart card data in Beijing and provide the excess commuting framework. Ma *et al.* [15] further analyze the individuals' commuting pattern considering multi-variables like departure time, travel distance, and the number of traveling days. Moreover, they provide an approach to identify commuters and transit commuting patterns.

#### 2) ACTIVITY PREDICTION

Trajectory data not only intuitively reflects human mobility pattern but also contains potential information on human activities in their daily life. Moreover, predicting human activities can benefit many transportation planners and companies.

For activity prediction, plenty of prediction methods are specialized and enhanced to apply to the specific cases. Researchers apply behavior-based algorithms to identify activity locations and their types from trajectory data by generating the simulated cell phone dataset in [22]. Besides the identification of locations and types of activities, Hasan and Ukkusuri [25] present a data-driven approach based on social networks to account for missing activities and provide an effective technique to infer disaggregate individual activity patterns.

Hidden Markov Model (HMM) is a stochastic model which is widely used to infer the activities. Widhalm *et al.* [24] propose a method which models the dependencies among activity types, trip scheduling, and land occupation types with a Relational Markov Network (RMN) to reveal activity behavioral pattern, and analyze relational signatures of activity time, duration, and land occupation. A similar methodology proposed by Han and Sohn [19] uses a Continuous Hidden Markov model (CHMM) coupled with the spatial and temporal information on trip chains from smart card data and land-use types to estimate the sequence of activities for each trip chain.

In recent years, methods of activities prediction tend to be specific in different areas especially urban transportation. Guo and Karimi [20] propose a novel methodology for the prediction of spatial-temporal activities, e.g., the inflow and outflow of people in neighborhoods/areas during certain time periods. The results show that prediction methodology achieves a high accuracy. Zheng and Zhou [21] propose a transportation prediction model which is not sensitive to time to examine the scaling laws of spatial visitation frequency and understand the influence of the built environment.

#### 3) ANOMALOUS EVENT DETECTION

Regular trajectory data own extreme values that deviate from other observations, that is abnormal data, which carries useful hidden information. In the traffic field, for example, there are growingly advanced sensing systems producing tremendous trajectory data for not only just analyzing but also detecting the anomalous events/behaviors violated traffic regulations, such as overspeed, drunk driving, traffic collision, hit-and-run, hijacking, and unexpected stops [31].

Before applying anomalous detection algorithms we have to figure out what are the main issues researchers try to deal with. Laxhammar and Falkman [27] address some problems of the detection of anomalous trajectories, e.g., the algorithm is designed for offline, insensitive to local sub-trajectory anomalies, and parameter-laden suffering from high alarm rates. To handle these issues, lots of algorithms based on analyzing normal trajectory patterns are proposed. Piciarelli *et al.* [28] present an approach, which is based on single-class Support Vector Machine (SVM) clustering, to detect anomalous events differing from typical patterns. Instead of the traditional similarity measurement in anomaly detection, Chang *et al.* [26] propose a novel abnormal trajectory detecting model, which is based on the periodic-behavior rule that moving objects usually follow similar entrances and exits in their paths. An anomaly model that integrates time, space, and spatial scale is proposed by Li *et al.* [29] using a joint representation of video appearances and it is beneficial for defining anomalies in various daily contexts, such as anomaly scale space.

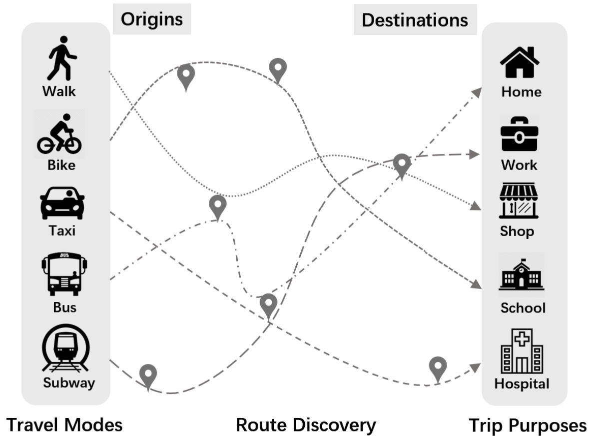


**FIGURE 2.** Travel patterns.

Different from approaches based on anomaly detection, approaches based on explicit event/behavior recognition with semantic interpretations are efficient. For example, a study [30] defines four types of extreme transit behaviors: early birds, night owls, tireless itinerants, and recurring itinerants. In addition, a framework is developed to identify the spatiotemporal patterns of four extreme transit behaviors, and the results are beneficial to public transportation management for providing public transit services to extreme transit riders.

Human behaviors own complex sequential transition regularities exhibited with time-dependent and high-order nature and multi-level periodicity. The variability of time and space in traffic environments brings more complexity to human mobility in traffic, that is, traffic patterns. Due to the complexity, traffic patterns get more interesting and attract attention of scholars in various fields. In addition to the three aspects above, traffic patterns present high academic and practical values in many sub-fields of urban planning, as well as other fields, like infectious disease research.

### *B. TRAVEL PATTERN*

Human travel pattern is a crucial part of our daily life. We display the travel pattern in Fig. 2. We select working, shopping, going home and others, as our travel purpose, and different purposes lead to different destinations. Because a pair of an origin and a destination can not generate a unique route in urban cities, the route discovery is also important. Besides, we choose a mode, such as walking, biking, driving or taking buses, to travel. We give a description of four aspects of travel pattern, trip purpose estimation, destination prediction, route discovery, and travel mode analysis.

#### 1) TRIP PURPOSE ESTIMATION

Trip purpose estimation is essential for urban planning, especially public transportation planning. Krizek [37] provides a comprehensive perspective on neighborhood services, trip purposes, and tour-based travel, especially the relationships between neighborhood accessibility and trip purposes. Besides, tour frequency or complexity is also considered in analyzing multi-purpose tours. Some researches on trip purpose estimation focus on influencing factors. Elldér [69] confirms that the influence of residential locations on daily travel distance is highly conditional on trip purposes which are classified by time-spatial constraints and hypothesized factors of personal choices. Gong *et al.* [33] point out a problem that it lacks semantic analysis of massive trajectory data is highly valuable for a range of applications and services, and provide a practical methodology to estimate the trip purposes of taxi passengers, meanwhile, it enriches the semantics of trajectory data.

A lot of researches are conducted for the different types of trajectory data. For the explicit trajectory data, Shen and Stopher [32] propose an effective process based on GPS data using additional trip information, such as activity duration and time of activities. Xiao *et al.* [36] propose a model for trip purpose combining artificial neural networks and particle swarm optimization from GPS data. Besides, the results of trip purposes are classified to evaluate the performance of this model.

The implicit trajectory data assumes that every travel within a trip sequence is based on the origin-destination matrix. For the sensor-based trajectory data, Lee and Hickman [34] introduce the practical use of AFC data to estimate trip purposes and address the problem of making transit users' trip purposes sense. As for the signal-based trajectory data, Alexander *et al.* [35] extract daily origin-destination trips by trip purposes using CDR. They also confirm that land use information can be used to infer trip purposes.

#### 2) DESTINATION PREDICTION

Destination prediction is essential for many emerging location-based services such as recommendations from travel agencies and personalized advertising based on destinations. In 2013, Xue *et al.* [38] point out a data sparsity problem in destination prediction by using common approaches like deriving the probability of a location being the destination based on historical trajectories and propose a novel method named Sub-Trajectory Synthesis (SubSyn), which decomposes historical trajectories into sub-trajectories and connect them into new trajectories. Later in 2015, Xue *et al.* [40] improve runtime efficiency and prediction accuracy of SubSyn to validate against a real-world and large-scale taxi trajectory dataset and discuss the data sparsity problem in detail. Kanno *et al.* [39] propose a real-time passenger location estimation method using CDRs and a crowdsourcing timetable.

#### 3) ROUTE DISCOVERY

Rapid urbanization and increasing travel demand make route discovery a critical task for urban planners. Route discovery is usually based on destination prediction. A practical methodology is proposed to predict a personal route over an entire day from candidate routes generated based on stay points detected from historical locations using CDRs and GPS data [42]. Apart from trajectory data, its related data is also quite useful in analyzing candidate routes. Toole *et al.* [43]

present a flexible and efficient system to estimate the travel demand and the routes most traveled using CDRs and GIS information, utilize massive raw data to analyze road network and build an interactive web visualization based on GIS information to explore road usage patterns and help make these results of route demands accessible to consumers and policymakers.

At the methodology level, the approaches of route discovery can be further extended by abstracting road networks to graph models. Tang *et al.* [44] propose a novel time-dependent graph model to predict the most likely route within a transportation network to analyze the uncertainty. Besides time-dependent model, space factors can enhance the availability of models. Yang *et al.* [45] present a space-time trajectory cube, which is organized by origin, destination, and time, and propose a framework which is used to compute the important information and extract the fine-grained experience of drivers from real trajectory data. Probabilistic models are also suitable for route discovery. Besides, a probabilistic model is developed to estimate passenger route choices in a complex metro network [46].

#### 4) TRAVEL MODE ANALYSIS

Travel modes of individuals are becoming increasingly varied with rapidly growing travel demand. Effect factors of travel modes are significant. Anable and Gatersleben [70] separately discuss the travel mode choice of work travel and leisure travel and the results reveal that the variables of identifying mode users' evaluations of performance on the aspects are vitally important to them.

A simple and general approach to estimate travel modes is based on travel time and speed with clear effect factors and specific targets. Bohte and Maat [47] propose a practical method by calculating the average and maximum speed to identify walking, biking or driving, and using GIS data like trackpoints to identify other modes of travel which have similar speed. There also exists another similar study. Yang *et al.* [49] propose an innovative method to estimate the travel information including travel modes, mode-changing time and location, and other attributes. It analyzes four kinds of machine learning algorithms for identifying walking, biking or driving, then proposes a critical point matching algorithm combining individuals GPS data with bus stop GIS information for better performance.

However, the speed-based methodologies above usually need trajectory related data like GIS to provide the extra information which makes it difficult to implement. Therefore, Zheng *et al.* [48] provide a comprehensive perspective about understanding travel modes based on GPS data, and propose a graph-based post-processing algorithm which is superior to normal post-processing. Jiang *et al.* [50] analyze three kinds of travel modes (taxi, subway, and bus), and point out that the displacements of taxi and bus trips follow the different exponential distribution, or the displacements of subway follow the gamma distribution. The results of this research are significant for travel mode estimation.

### C. OTHERS

In this section, we give a detailed introduction of applications of trajectory data from three aspects which are not assigned to the above two categories, urban function, time inference, and environments.

#### 1) URBAN FUNCTIONS

In urban function field, the applications of trajectory data are of great concern to deal with urban puzzles, especially for the government. Understanding urban function deeply can help urban planners to make reasonable policies and to improve traffic conditions. However, the discovery of urban functions is challenging due to the massive urban buildings and complex influencing factors. Zhong *et al.* [53] propose a method to estimate urban functions at the building level using smart card data. A probabilistic framework, which is based on daily activities can reveal how people use urban space in reality. Then it is established on the basis of relations among trips, stops, and buildings. Hu *et al.* [55] propose a framework to understand the urban AOI using DBSCAN clustering algorithm, which extracts the distinctive texture information from geotag photos and analyze the spatial dynamics as well as the insights derived from urban AOI. Sun *et al.* [51] provide a practical community detection method and a comprehensive analysis of community structure in urban traffic zones in network science. The results reveal that traffic communities are also related to the travel demand distribution. As for influencing factors of urban function analysis, Zhu *et al.* [52] point out that streets, as the basic elements of the urban cities, are significant for depicting urban functions and they discuss the differences or relationship of the linear street units and traditional areal units.

#### 2) TIME INFERENCE

Time inference is particularly significant for improving urban traffic operations, assessing the efficiency and performance of transportation networks and management systems. It can not only benefit passengers from accurate time information, which brings great convenience to individuals' daily life, but also can offer opportunities for travel companies to make choices for users by minimizing overall travel time. Mori *et al.* [58] provide a comprehensive perspective of time inference and differentiate travel time estimation from travel time prediction. However, several problems bring resistance to the analysis of time inference. Data sparsity, an optimal combination of routes to estimate time is hard to find the query efficiency [57]. For the data sparsity problem, Sanaullah *et al.* [59] identify a series of influencing factors, e.g., GPS sampling frequency, vehicle penetration rate, and length of the time window, in order to develop a practical method for travel time estimation. Zhan *et al.* [56] propose a data-driven model using large-scale taxi data which lacks the information of actual paths taken by the drivers, and use an embedded MNL model to compute the probability of a given path in the constructed candidate paths. Signal-based trajectory data can also be utilized to estimate time of short distance travel. Abedi *et al.* [60] propose a method to estimate the travel time of walkers, runners, and cyclists in the real scenario by analyzing Wi-Fi and Bluetooth data.

#### 3) ENVIRONMENT MONITORING

The expending of urbanization leads to the increase of exhaust gas emissions. Researches on environmental pollution based on trajectory data provide critical insights for a series of future social and environmental implications on urban transportation management. Fuel consumptions and emissions of vehicles can be predicted using explanatory variables like instantaneous speed and acceleration [61]. An extended Kalman filter algorithm, which contains various elements including the dynamic model, measurement equations and the formulation of the EKF, is designed for the navigational function of a real-time vehicle performance and emissions monitoring [62]. Using Aggregated Tracking of GPS-equipped Vehicles data, Chen *et al.* [64] present a methodology to analyze the traffic-related air pollution emissions with multiple traffic-related variables. It can estimate ambient pollutant gases by using a non-linear model that includes basic dispersion properties and then can validate them. Considering the impacts of traffic congestion, a novel framework to quantify air pollution is presented to combine the localized air quality records from AQS monitors and the annual pollution reported by the NEI [66]. Luo *et al.* [65] present a detailed energy consumption and emissions analysis in Shanghai, China, and explore related spatial and temporal features of emissions which present a distribution of dual-core cyclic structure.

## IV. SERVICES OF TRAJECTORY DATA

Applications of trajectory data are mainly to solve a specific problem by a single method. Such problems are abstracted from real life, academically and idealized. In this section, we focus on the services of trajectory data. Different from applications, services are closer to real life. Moreover, a variety of techniques and methods are integrated to solve practical problems. It can consider issues from the perspective of users. So next we provide a comprehensive introduction of the services of trajectory data from the perspective of two types of users, such as the government and the commercial organizations.

### A. TRANSPORTATION ADMINISTRATION

The main purpose of the services provided by the urban administrators is to meet the needs of residents rather than gain income. Thus, the services of urban management mainly focus on three aspects: urban planning, public services, and real-time monitoring.

#### 1) URBAN PLANNING

With the explosive growth of urban population and vehicles' number, various urban problems being solved emerge constantly, such as traffic congestion and exhaust emissions. In-depth understanding of cities through trajectory data of residents or vehicles is an effective means of solving urban problems.

Urban structure planning is a basic and crucial issue in urban planning, such as land use and transportation network design. Liu *et al.* [71] reveal the temporal variations of urban land use such as features and densities, and characterize traffic source-sink areas by land use features. A two-stage algorithm is presented to plan and design the smart railway trajectory on GIS system considering the distance factor, the construction cost, the altitude factor and the comfortability [72]. By inferring the transportation demand under monitoring territory, Nanni *et al.* [73] present a practical process for transportation planning. A node place model is proposed to identify the most effective transportation and land use dynamics in station areas by analyzing the spatial relationship with other transportation and land uses [74]. Using a large sample of user location data, Pinelli *et al.* [75] present a data-driven methodology to design the transit network with spatial resolutions at the level of cell tower locations, which provides a service that would be effective for all citizens. Su *et al.* [76] propose an approach for intra-urban food service planning by analyzing healthy food accessibility of urban communities based on the travel time calculations considering the variables.

Urban infrastructure planning is a long-term process which requires analysis and consensus of multiple demands and complex goals. By analyzing the charging demand using the taxi trajectory data, Shahraki *et al.* [77] propose a novel model to select the optimal location of public charging stations for maximizing the vehicle-miles-traveled. A practical methodology is presented for designing large-scale infrastructure, like Electrical Road Systems, to help urban planners and decision makers electrify urban transport systems [78].

In a word, based on the trajectory data, urban planning is more reasonable than the planning done by experience. The government can provide residents with a more comfortable travel environment.

#### 2) PUBLIC SERVICE OPTIMIZATION

Public service is a significant part of citizen daily life especially public transportation services, like buses and taxis. However, a series of troubles need to be solved to facilitate public transportation services, and transportation managers are faced with massive challenges about unstable traffic conditions, e.g., the mismatch of taxi services and mobility demand, poor on-time performance and bunching problems of bus services.

A methodology is presented to deduce passengers' route choices, which is important for analyzing passenger services in terms of travel time [79]. A reliable method is presented to analyze individuals' mobility demand and transportation network services, and accurately identify areas with serious mismatch problems between travel demand and transportation services [80]. Farber *et al.* [81] compare the mobility demand and the transit service supply based on travel times, and characterize the transit supply using a three-dimensional transit travel time cube. The results confirm the theory which is "more marginalized groups demand travel between locations at times of the day that are poorly served by transit". A practical methodology for real-time trajectory monitoring is presented to improve taxi services [82]. To quantify the demand dynamics for supporting optimal taxi services strategies, Kourti *et al.* [83] apply dynamic clustering and "heatmap" analyzing to a complex and realistic GPS dataset to preliminarily identify the alternative Taxi-Services Strategies in urban cities.

We can trust that in the future, people can not travel without the help of trajectory data. Trajectory data is a valuable asset in terms of route planning and taxi services [84].

#### 3) REAL-TIME MONITORING

Trajectory data not only provides urban dynamics of traffic control but also monitors the spreading of infectious diseases even human contact pattern. A practical method is proposed to analyze and evaluate the spatial spreading of infectious diseases in both sort-scale commuting and long-range airline traffic [85]. Zhu *et al.* [86] propose a practical visualization technique to discover and analyze the time and the position that people from different locations move into the same places and make contact. A real-time urban monitoring system is developed to obtain the instantaneous location of buses and taxis and collect the traffic voice and data to describe the traffic statues [87]. Based on the systematic analysis of traffic requirements, a comprehensive and flexible architecture is proposed to inform real-time traffic control logic and accelerate transportation operation [88]. Zhu *et al.* [86] propose a framework to instantly detect suspicious companion vehicles when they pass through monitoring systems in city.

In short, real-time monitoring is an extremely accurate application, which needs dynamic real-time trajectory data. The government can use trajectory data to set up the monitoring and precaution system.

### B. COMMERCIAL SERVICES

Commercial organizations provide competitive and attractive services to maintain a customer base to make profits. For example, compared with taxis, Didi and Uber can offer a cheap and efficient mode to take short trips. The personalized recommendation is an effective way to improve customers' satisfaction with services offered by commercial organizations. Trajectory recommendation problems can be divided into three categories: activity-based recommendation, GPS-based recommendation and hybrid recommendation [89]. A comprehensive perspective of recommendation services based on network-based trajectory data summarizes the methodologies employed to generate a recommendation [90]. Leung *et al.* [91] propose a framework called collaborative location recommendation by dividing users to different classes to address two problems of clustering user location matrix: the difficulty of selecting the optimal candidate location from numerous similar locations, and the time consumption and efficiency of clustering massive location matrix.

In addition, a time-location-relationship combined taxi service recommendation model is proposed to solve the empty carrying phenomenon and improve the taxi drivers' profits and satisfy the passengers traveling need [92]. Hwang *et al.* [93] analyze the vital factors which infect the quality of recommendation systems, e.g., distance to the recommended location, waiting time for the next passengers, and expected fare for the trip. And they discuss the influence of driver experiences or preferences which are key factors in detail. In the waiting time aspect, Xu *et al.* [94] propose a practical system called Taxi-hunting Recommendation System to pick up passengers efficiently with a short waiting time. As for distance, Yang *et al.* [95] propose Adaptive Shortest Expected Cruising Route algorithm and implement a novel data structure to improve the performance of recommendation systems.

Different from the government service, the commercial service aims at gaining profit. With the help of trajectory data, companies and enterprises can gain the maximum benefit with minimal resource consumption, and provide the best service to users.

## V. CHALLENGES

In previous sections, we survey several key issues of trajectory data, including the classifications, the applications and the services of trajectory data. However, there are still many challenges which are representative of critical directions. We give some challenges that seem promising for further research in this section.

### A. PRIVACY PROTECTION

The ubiquity of mobile devices and the improvement of coverage and accuracy for GPS receivers make the privacy issue a difficult problem. Considering the worst case of tracking bound and achieving significant data accuracy, a time-confusion criterion and an uncertainty-aware path cloaking algorithm are proposed to hide location samples to overcome the challenges of failing to provide privacy guarantees in a low-density area [96]. Hoh *et al.* [97] develop a system based on virtual trip lines which can indicate where vehicles provide location information and utilize associated cloaking techniques to avoid the particularly privacy sensitive locations. Hwang *et al.* [98] present a novel time-obfuscated algorithm for trajectory privacy protection which combines ambient conditions to cloak location information and is able to prevent malicious LBS reconstructing trajectory. A new trajectory privacy-preserving framework is proposed by Gao *et al.* [99] and this framework improves the mix-zones model considering the time factor, and evaluates the effectiveness on the basis of information entropy with previous models. Theodorakopoulos *et al.* [100] present a method to achieve location privacy-preserving mechanisms considering the predictability of passengers' whereabouts and sequential correlation. This method is the first to protect the privacy of transitions between locations. Considering the density of trajectories and diversity of users, a methodology is proposed to produce a privacy-preserving heat map with user diversity [101].

At present, none of the existing privacy-preserving task allocation mechanisms can provide personalized location protection considering different protection demands of workers. Wang *et al.* [102] propose a personalized privacy-preserving task allocation framework for mobile crowdsensing that can allocate tasks effectively while providing personalized location privacy protection. Besides, Wang *et al.* [103] study the problem of real-time crowdsourced statistical data publishing with strong privacy protection under an untrusted server. They propose a novel distributed agent-based privacy-preserving framework that introduces a new level of multiple agents between the users and the untrusted server.

### B. HUMAN MOBILITY CAUSALITY

There are lots of studies about the changes of human mobility and the influence factors. Araki *et al.* [11] conduct a basic analysis of the impacts of seasonal factors on human mobility by using random forest model and GSP data. The global or local spatial variation and correlation of urban human mobility are revealed by Li *et al.* [12]. Briand *et al.* [14] analyze year-to-year changes in passengers' mobility pattern in public transport using smart card data. Feng *et al.* [104] provide multi-variate analysis about mode choice for commuting and shopping-leisure trips and daily travel distance using the data of 2008-2011 in Nanjing. Co-occurrence pattern is an interesting human mobility pattern, which means people from different places visit the same urban regions at same time intervals and can be mined from GPS data and CDR [105]. However, while a lot of influence factors of human mobility, it is still hard to describe and predict the human mobility. In the future, researchers need to develop standard rubrics and benchmarks for evaluating these different factors.

### C. EMISSION REDUCTION

With the development of motor vehicles, the estimating total emission and local contributions of carbon emission are also worth to be studied. Nowadays, although researchers have present many methods to decrease the carbon emission, there is still a long way to get satisfactory results.

Novel detection models can be built by combining sensor-based trajectory data and related data. Wang *et al.* [67] develop a GIS-based software to identify potential sources from long-term air pollution utilizing various trajectory statistical analysis methods. An online model is proposed for air pollution monitoring using the GPRS public network and adopts the data of pollutant gases such as CO, $NO_2$, and $SO_2$, which are collected by city buses, and transmitted through Google Maps to make the data available [63].

As for local contributions of carbon emission, Liu and Wang [68] present a multiproxy allocation system to identify and map local contributions of carbon emissions from urban motor and metro transports, meanwhile they use a top-down approach to allocated local carbon-emitting quantities from total, per capita, and per unit perspectives so as to reveal the spatial differences between these perspectives. This study provides insights into the effective and reasonable allocation of transport carbon emissions on local.

## VI. CONCLUSION

In this paper, we provide a literature review on categories, applications, service and privacy protection of trajectory data. At first, we introduce the concepts of trajectory data and classify the trajectory data by the data forms into explicit trajectory data and implicit trajectory data. Then we offer a systematic review of applications of various trajectory data. Many issues still exist in the applications of trajectory data, e.g., data sparseness, and the efficiency of processing and querying massive trajectory data. However, the development of data acquisition technologies and methodologies of trajectory data mining can solve these problems to some extent in the future. Additionally, we provide a comprehensive perspective on the services of trajectory data that is discussed in detail both from the aspects of the government and commercial organizations. Finally, the significant challenges in the trajectory data analysis are discussed in detail as well.

## REFERENCES

[1] M. Chen, S. Mao, and Y. Liu, "Big data: A survey," *Mobile Netw. Appl.*, vol. 19, no. 2, pp. 171–209, Apr. 2014.

[2] Y. Zheng, "Trajectory data mining: An overview," *ACM Trans. Intell. Syst. Technol.*, vol. 6, no. 3, May 2015, Art. no. 29.

[3] J. D. Mazimpaka and S. Timpf, "Trajectory data mining: A review of methods and applications," *J. Spatial Inf. Sci.*, no. 13, pp. 61–99, Dec. 2016.

[4] X. J. Kong *et al.*, "Mobility dataset generation for vehicular social networks based on floating car data," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 3874–3886, May 2018.

[5] M.-P. Pelletier, M. Trépanier, and C. Morency, "Smart card data use in public transit: A literature review," *Transp. Res. C, Emerg. Technol.*, vol. 19, no. 4, pp. 557–568, 2011.

[6] F. Giannotti, L. Pappalardo, D. Pedreschi, and D. Wang, "A complexity science perspective on human mobility," in *Mobility Data*, C. Renso, S. Spaccapietra, and E. Zimányi, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2013, pp. 297–314.

[7] 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.

[8] K. K. Jahromi, M. Zignani, S. Gaito, and G. P. Rossi, "Simulating human mobility patterns in urban areas," *Simul. Model. Pract. Theory*, vol. 62, pp. 137–156, Mar. 2016.

[9] S.-L. Shaw, M.-H. Tsou, and X. Ye, "Editorial: Human dynamics in the mobile and big data era," *Int. J. Geograph. Inf. Sci.*, vol. 30, no. 9, pp. 1687–1693, Sep. 2016.

[10] C.-Z. Yao and J.-N. Lin, "A study of human mobility behavior dynamics: A perspective of a single vehicle with taxi," *Transp. Res. A, Policy Pract.*, vol. 87, pp. 51–58, May 2016.

[11] M. Araki, R. Kanamori, L. Gong, and T. Morikawa, "Impacts of seasonal factors on travel behavior: Basic analysis of GPS trajectory data for 8 months," in *Serviceology for Smart Service System*, Y. Sawatani, J. Spohrer, S. Kwan, and T. Takenaka, Eds. Tokyo, Japan: Springer, 2017, pp. 377–384.

[12] X. Li, W. Tu, S. Shaoqing, Y. Yang, N. Luo, and Q. Li, "Revealing spatial variation and correlation of urban travels from big trajectory data," *ISPRS-Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. XLII-2/W7, pp. 53–57, Sep. 2017.

[13] J. Zhou, E. Murphy, and Y. Long, "Commuting efficiency in the Beijing metropolitan area: An exploration combining smartcard and travel survey data," *J. Transport Geogr.*, vol. 41, pp. 175–183, Dec. 2014.

[14] A.-S. Briand, E. Côme, M. Trépanier, and L. Oukhellou, "Analyzing year-to-year changes in public transport passenger behaviour using smart card data," *Transp. Res. C, Emerg. Technol.*, vol. 79, pp. 274–289, Jun. 2017.

[15] X. Ma, C. Liu, H. Wen, Y. Wang, and Y. Wu, "Understanding commuting patterns using transit smart card data," *J. Transp. Geogr.*, vol. 58, pp. 135–145, Jan. 2017.

[16] T. Prentow, A. Thom, H. Blunck, and J. Vahrenhold, "Making sense of trajectory data in indoor spaces," in *Proc. 16th IEEE Int. Conf. Mobile Data Manage.*, vol. 1, Jun. 2015, pp. 116–121.

[17] A. N. Larijani, A.-M. Olteanu-Raimond, J. Perret, M. Brédif, and C. Ziemlicki, "Investigating the mobile phone data to estimate the origin destination flow and analysis; case study: Paris region," *Transp. Res. Procedia*, vol. 6, pp. 64–78, Apr. 2015.

[18] Y. Qiao, Y. Cheng, J. Yang, J. Liu, and N. Kato, "A mobility analytical framework for big mobile data in densely populated area," *IEEE Trans. Veh. Technol.*, vol. 66, no. 2, pp. 1443–1455, Feb. 2017.

[19] G. Han and K. Sohn, "Activity imputation for trip-chains elicited from smart-card data using a continuous hidden Markov model," *Transp. Res. B, Methodol.*, vol. 83, pp. 121–135, Jan. 2016.

[20] Q. Guo and H. A. Karimi, "A novel methodology for prediction of spatial-temporal activities using latent features," *Comput., Environ. Urban Syst.*, vol. 62, pp. 74–85, Mar. 2017.

[21] Z. Zheng and S. Zhou, "Scaling laws of spatial visitation frequency: Applications for trip frequency prediction," *Comput., Environ. Urban Syst.*, vol. 64, pp. 332–343, Jul. 2017.

[22] C. Chen, L. Bian, and J. Ma, "From traces to trajectories: How well can we guess activity locations from mobile phone traces?" *Transp. Res. C, Emerg. Technol.*, vol. 46, pp. 326–337, Sep. 2014.

[23] M. Shoaib, S. Bosch, O. D. Incel, H. Scholten, and P. J. M. Havinga, "A survey of online activity recognition using mobile phones," *Sensors*, vol. 15, no. 1, pp. 2059–2085, Jan. 2015.

[24] P. Widhalm, Y. Yang, M. Ulm, S. Athavale, and M. C. González, "Discovering urban activity patterns in cell phone data," *Transportation*, vol. 42, no. 4, pp. 597–623, Jul. 2015.

[25] S. Hasan and S. V. Ukkusuri, "Urban activity pattern classification using topic models from online geo-location data," *Transp. Res. C, Emerg. Technol.*, vol. 44, pp. 363–381, Jul. 2014.

[26] C.-W. Chang, M.-H. Yang, C.-C. Li, and K.-T. Chuang, "Toward mining anomalous behavior from big moving trajectories in surveillance video," in *Proc. IEEE Int. Conf. Autom. Sci. Eng. (CASE)*, Aug. 2014, pp. 1121–1126.

[27] R. Laxhammar and G. Falkman, "Inductive conformal anomaly detection for sequential detection of anomalous sub-trajectories," *Ann. Math. Artif. Intell.*, vol. 74, nos. 1–2, pp. 67–94, Jun. 2015.

[28] C. Piciarelli, C. Micheloni, and G. L. Foresti, "Trajectory-based anomalous event detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 11, pp. 1544–1554, Nov. 2008.

[29] W. Li, V. Mahadevan, and N. Vasconcelos, "Anomaly detection and localization in crowded scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 1, pp. 18–32, Jan. 2014.

[30] Y. Long, X. Liu, J. Zhou, and Y. Chai, "Early birds, night owls, and tireless/recurring itinerants: An exploratory analysis of extreme transit behaviors in Beijing, China," *Habitat Int.*, vol. 57, pp. 223–232, Oct. 2016.

[31] X. J. Kong, X. Song, F. Xia, H. Guo, J. Wang, and A. Tolba, "LoTAD: Long-term traffic anomaly detection based on crowdsourced bus trajectory data," *World Wide Web*, vol. 21, no. 3, pp. 825–847, 2017.

[32] L. Shen and P. R. Stopher, "A process for trip purpose imputation from global positioning system data," *Transp. Res. C, Emerg. Technol.*, vol. 36, pp. 261–267, Nov. 2013.

[33] L. Gong, X. Liu, L. Wu, and Y. Liu, "Inferring trip purposes and uncovering travel patterns from taxi trajectory data," *Cartogr. Geograph. Inf. Sci.*, vol. 43, no. 2, pp. 103–114, Mar. 2016.

[34] S. G. Lee and M. Hickman, "Trip purpose inference using automated fare collection data," *Public Transp.*, vol. 6, no. 1, pp. 1–20, Apr. 2014.

[35] L. Alexander, S. Jiang, M. Murga, and M. C. González, "Origin–destination trips by purpose and time of day inferred from mobile phone data," *Transp. Res. C, Emerg. Technol.*, vol. 58, pp. 240–250, Sep. 2015.

[36] G. Xiao, Z. Juan, and C. Zhang, "Detecting trip purposes from smartphone-based travel surveys with artificial neural networks and particle swarm optimization," *Transp. Res. C, Emerg. Technol.*, vol. 71, pp. 447–463, Oct. 2016.

[37] K. J. Krizek, "Neighborhood services, trip purpose, and tour-based travel," *Transportation*, vol. 30, no. 4, pp. 387–410, Nov. 2003.

[38] 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. IEEE 29th Int. Conf. Data Eng. (ICDE)*, Apr. 2013, pp. 254–265.

[39] T. Kanno, H. Kanasugi, Y. Sekimoto, and R. Shibasaki, "Real-time passenger location estimation using CDRs and train objects generated from crowdsourced timetables," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput., ACM Int. Symp. Wearable Comput.*, 2015, pp. 1197–1205.

[40] A. Y. Xue, J. Qi, X. Xie, R. Zhang, J. Huang, and Y. Li, "Solving the data sparsity problem in destination prediction," *VLDB J.*, vol. 24, no. 2, pp. 219–243, Apr. 2015.

[41] O. Burkhard, R. Ahas, E. Saluveer, and R. Weibel, "Extracting regular mobility patterns from sparse CDR data without *a priori* assumptions," *J. Location Based Services*, vol. 11, pp. 1–20, Jun. 2017.

[42] H. Kanasugi, Y. Sekimoto, M. Kurokawa, T. Watanabe, S. Muramatsu, and R. Shibasaki, "Spatiotemporal route estimation consistent with human mobility using cellular network data," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops (PERCOM Workshops)*, Mar. 2013, pp. 267–272.

[43] J. L. Toole, S. Colak, B. Sturt, L. P. Alexander, A. Evsukoff, and M. C. González, "The path most traveled: Travel demand estimation using big data resources," *Transp. Res. C, Emerg. Technol.*, vol. 58, pp. 162–177, Sep. 2015.

[44] J. Tang, Y. Song, H. J. Miller, and X. Zhou, "Estimating the most likely space–time paths, dwell times and path uncertainties from vehicle trajectory data: A time geographic method," *Transp. Res. C, Emerg. Technol.*, vol. 66, pp. 176–194, May 2016.

[45] L. Yang, M.-P. Kwan, X. Pan, B. Wan, and S. Zhou, "Scalable space-time trajectory cube for path-finding: A study using big taxi trajectory data," *Transp. Res. B, Methodol.*, vol. 101, pp. 1–27, Jul. 2017.

[46] J. Zhao *et al.*, "Estimation of passenger route choice pattern using smart card data for complex metro systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 4, pp. 790–801, Apr. 2017.

[47] W. Bohte and K. Maat, "Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in The Netherlands," *Transp. Res. C, Emerg. Technol.*, vol. 17, no. 3, pp. 285–297, Jun. 2009.

[48] Y. Zheng, Y. Chen, Q. Li, X. Xie, and W.-Y. Ma, "Understanding transportation modes based on GPS data for Web applications," *ACM Trans. Web*, vol. 4, no. 1, p. 1, 2010.

[49] F. Yang, Z. Yao, Y. Cheng, B. Ran, and D. Yang, "Multimode trip information detection using personal trajectory data," *J. Intell. Transp. Syst.*, vol. 20, no. 5, pp. 449–460, Sep. 2016.

[50] S. Jiang, W. Guan, W. Zhang, X. Chen, and L. Yang, "Human mobility in space from three modes of public transportation," *Phys. A, Stat. Mech. Appl.*, vol. 483, pp. 227–238, Oct. 2017.

[51] L. Sun, X. Ling, K. He, and Q. Tan, "Community structure in traffic zones based on travel demand," *Phys. A, Stat. Mech. Appl.*, vol. 457, pp. 356–363, Sep. 2016.

[52] D. Zhu, N. Wang, Y. Liu, and L. Wu, "Street as a big geo-data assembly and analysis unit in urban studies: A case study using Beijing taxi data," *Appl. Geogr.*, vol. 86, pp. 152–164, Sep. 2017.

[53] C. Zhong, X. Huang, S. M. Arisona, G. Schmitt, and M. Batty, "Inferring building functions from a probabilistic model using public transportation data," *Comput., Environ. Urban Syst.*, vol. 48, pp. 124–137, Nov. 2014.

[54] Y. Gong, Y. Lin, and Z. Duan, "Exploring the spatiotemporal structure of dynamic urban space using metro smart card records," *Comput., Environ. Urban Syst.*, vol. 64, pp. 169–183, Jul. 2017.

[55] Y. Hu, S. Gao, K. Janowicz, B. Yu, W. Li, and S. Prasad, "Extracting and understanding urban areas of interest using geotagged photos," *Comput., Environ. Urban Syst.*, vol. 54, pp. 240–254, Nov. 2015.

[56] X. Zhan, S. Hasan, S. V. Ukkusuri, and C. Kamga, "Urban link travel time estimation using large-scale taxi data with partial information," *Transp. Res. C, Emerg. Technol.*, vol. 33, pp. 37–49, Aug. 2013.

[57] Y. Wang, Y. Zheng, and Y. Xue, "Travel time estimation of a path using sparse trajectories," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2014, pp. 25–34.

[58] U. Mori, A. Mendiburu, M. Álvarez, and J. A. Lozano, "A review of travel time estimation and forecasting for advanced traveller information systems," *Transportmetrica A, Transport Sci.*, vol. 11, no. 2, pp. 119–157, Feb. 2015.

[59] I. Sanaullah, M. Quddus, and M. Enoch, "Developing travel time estimation methods using sparse GPS data," *J. Intell. Transp. Syst.*, vol. 20, no. 6, pp. 532–544, Nov. 2016.

[60] N. Abedi, A. Bhaskar, E. Chung, and M. Miska, "Assessment of antenna characteristic effects on pedestrian and cyclists travel-time estimation based on Bluetooth and WiFi MAC addresses," *Transp. Res. C, Emerg. Technol.*, vol. 60, pp. 124–141, Nov. 2015.

[61] K. Ahn, H. Rakha, A. Trani, and M. van Aerde, "Estimating vehicle fuel consumption and emissions based on instantaneous speed and acceleration levels," *J. Transp. Eng.*, vol. 128, no. 2, pp. 182–190, Mar. 2002.

[62] L. Zhao, W. Y. Ochieng, M. A. Quddus, and R. B. Noland, "An extended Kalman filter algorithm for integrating GPS and low cost dead reckoning system data for vehicle performance and emissions monitoring," *J. Navigat.*, vol. 56, no. 2, pp. 257–275, May 2003.

[63] A. R. Al-Ali, I. Zualkernan, and F. Aloul, "A mobile GPRS-sensors array for air pollution monitoring," *IEEE Sensors J.*, vol. 10, no. 10, pp. 1666–1671, Oct. 2010.

[64] S. Chen, S. Bekhor, Yuval, and D. M. Broday, "Aggregated GPS tracking of vehicles and its use as a proxy of traffic-related air pollution emissions," *Atmos. Environ.*, vol. 142, pp. 351–359, Oct. 2016.

[65] X. Luo *et al.*, "Analysis on spatial-temporal features of taxis' emissions from big data informed travel patterns: A case of Shanghai, China," *J. Cleaner Prod.*, vol. 142, pp. 926–935, Jan. 2017.

[66] C. K. Gately, L. R. Hutyra, S. Peterson, and I. S. Wing, "Urban emissions hotspots: Quantifying vehicle congestion and air pollution using mobile phone GPS data," *Environ. Pollut.*, vol. 229, pp. 496–504, Oct. 2017.

[67] Y. Q. Wang, X. Y. Zhang, and R. R. Draxler, "TrajStat: GIS-based software that uses various trajectory statistical analysis methods to identify potential sources from long-term air pollution measurement data," *Environ. Model. Softw.*, vol. 24, no. 8, pp. 938–939, Aug. 2009.

[68] C. Liu and T. Wang, "Identifying and mapping local contributions of carbon emissions from urban motor and metro transports: A weighted multiproxy allocating approach," *Comput., Environ. Urban Syst.*, vol. 64, pp. 132–143, Jul. 2017.

[69] E. Elldér, "Residential location and daily travel distances: The influence of trip purpose," *J. Transport Geogr.*, vol. 34, pp. 121–130, Jan. 2014.

[70] J. Anable and B. Gatersleben, "All work and no play? The role of instrumental and affective factors in work and leisure journeys by different travel modes," *Transp. Res. A, Policy Pract.*, vol. 39, nos. 2–3, pp. 163–181, Feb./Mar. 2005.

[71] Y. Liu, F. Wang, Y. Xiao, and S. Gao, "Urban land uses and traffic 'source-sink areas': Evidence from GPS-enabled taxi data in Shanghai," *Landscape Urban Planning*, vol. 106, no. 1, pp. 73–87, May 2012.

[72] J. Liu and T. Shi, "A new two-stages PSO algorithm for railway trajectory planning on GIS system," in *Proc. IEEE 17th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2014, pp. 2768–2773.

[73] M. Nanni *et al.*, "Transportation planning based on GSM traces: A case study on Ivory Coast," in *Citizen in Sensor Networks* (Lecture Notes in Computer Science). Cham, Switzerland: Springer, 2014, pp. 15–25.

[74] S. Monajem and F. E. Nosratian, "The evaluation of the spatial integration of station areas via the node place model; an application to subway station areas in Tehran," *Transp. Res. D, Transport Environ.*, vol. 40, pp. 14–27, Oct. 2015.

[75] F. Pinelli, R. Nair, F. Calabrese, M. Berlingerio, G. Di Lorenzo, and M. L. Sbodio, "Data-driven transit network design from mobile phone trajectories," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 6, pp. 1724–1733, Jun. 2016.

[76] S. Su, Z. Li, M. Xu, Z. Cai, and M. Weng, "A geo-big data approach to intra-urban food deserts: Transit-varying accessibility, social inequalities, and implications for urban planning," *Habitat Int.*, vol. 64, pp. 22–40, Jun. 2017.

[77] N. Shahraki, H. Cai, M. Turkay, and M. Xu, "Optimal locations of electric public charging stations using real world vehicle travel patterns," *Transp. Res. D, Transp. Environ.*, vol. 41, pp. 165–176, Dec. 2015.

[78] V. M. Shreenath and S. Meijer, "Spatial big data for designing large scale infrastructure: A case-study of electrical road systems," in *Proc. 3rd IEEE/ACM Int. Conf. Big Data Comput., Appl. Technol.*, 2016, pp. 143–148.

[79] E. van der Hurk, L. Kroon, G. Maróti, and P. Vervest, "Deduction of passengers' route choices from smart card data," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 1, pp. 430–440, Feb. 2015.

[80] J. Cui, F. Liu, J. Hu, D. Janssens, G. Wets, and M. Cools, "Identifying mismatch between urban travel demand and transport network services using GPS data: A case study in the fast growing Chinese city of Harbin," *Neurocomputing*, vol. 181, pp. 4–18, Mar. 2016.

[81] S. Farber, B. Ritter, and L. Fu, "Space–time mismatch between transit service and observed travel patterns in the Wasatch Front, Utah: A social equity perspective," *Travel Behav. Soc.*, vol. 4, pp. 40–48, May 2016.

[82] Z. Zhou *et al.*, "A method for real-time trajectory monitoring to improve taxi service using GPS big data," *Inf. Manage.*, vol. 53, no. 8, pp. 964–977, Dec. 2016.

[83] E. Kourti, C. Christodoulou, L. Dimitriou, S. Christodoulou, and C. Antoniou, "Quantifying demand dynamics for supporting optimal taxi services strategies," *Transp. Res. Procedia*, vol. 22, pp. 675–684, May 2017.

[84] X. Kong, M. Li, T. Tang, K. Tian, L. Moreira-Matias, and F. Xia, "Shared subway shuttle bus route planning based on transport data analytics," *IEEE Trans. Autom. Sci. Eng.*, vol. 15, pp. 1507–1520, Sep. 2018, doi: 10.1109/TASE.2018.2865494.

[85] D. Balcan, V. Colizza, B. Gonçalves, H. Hu, J. J. Ramasco, and A. Vespignani, "Multiscale mobility networks and the spatial spreading of infectious diseases," *Proc. Nat. Acad. Sci. USA*, vol. 106, no. 51, pp. 21484–21489, 2009.

[86] M. Zhu, C. Liu, J. Wang, X. Wang, and Y. Han, "Instant discovery of moment companion vehicles from big streaming traffic data," in *Proc. Int. Conf. Cloud Comput. Big Data (CCBD)*, Nov. 2015, pp. 73–80.

[87] F. Calabrese, M. Colonna, P. Lovisolo, D. Parata, and C. Ratti, "Real-time urban monitoring using cell phones: A case study in Rome," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 1, pp. 141–151, Mar. 2011.

[88] S. Amini, I. Gerostathopoulos, and C. Prehofer, "Big data analytics architecture for real-time traffic control," in *Proc. 5th IEEE Int. Conf. Models Technol. Intell. Transp. Syst. (MT-ITS)*, Jun. 2017, pp. 710–715.

[89] P. Yin, M. Ye, W.-C. Lee, and Z. Li, "Mining GPS data for trajectory recommendation," in *Proc. Pacific-Asia Conf. Knowl. Discovery Data Mining*. Cham, Switzerland: Springer, 2014, pp. 50–61.

[90] 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, Jul. 2015.

[91] K. W.-T. Leung, D. L. Lee, and W.-C. Lee, "CLR: A collaborative location recommendation framework based on co-clustering," in *Proc. 34th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2011, pp. 305–314.

[92] X. Kong, F. Xia, J. Wang, A. Rahim, and S. K. Das, "Time-location-relationship combined service recommendation based on taxi trajectory data," *IEEE Trans. Ind. Informat.*, vol. 13, no. 3, pp. 1202–1212, Jun. 2017.

[93] R.-H. Hwang, Y.-L. Hsueh, and Y.-T. Chen, "An effective taxi recommender system based on a spatio-temporal factor analysis model," *Inf. Sci.*, vol. 314, pp. 28–40, Sep. 2015.

[94] X. Xu, J. Zhou, Y. Liu, Z. Xu, and X. Zhao, "Taxi-RS: Taxi-hunting recommendation system based on taxi GPS data," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 4, pp. 1716–1727, Aug. 2015.

[95] W. Yang, X. Wang, S. M. Rahimi, and J. Luo, "Recommending profitable taxi travel routes based on big taxi trajectories data," in *Advances in Knowledge Discovery and Data Mining*, vol. 9078, T. Cao, E. P. Lim, Z. H. Zhou, T. B. Ho, D. Cheung, and H. Motoda, Eds. Cham, Switzerland: Springer, 2015, pp. 370–382.

[96] B. Hoh, M. Gruteser, H. Xiong, and A. Alrabady, "Preserving privacy in GPS traces via uncertainty-aware path cloaking," in *Proc. 14th ACM Conf. Comput. Commun. Secur.*, 2007, pp. 161–171.

[97] B. Hoh *et al.*, "Virtual trip lines for distributed privacy-preserving traffic monitoring," in *Proc. 6th Int. Conf. Mobile Syst., Appl., Services*, 2008, pp. 15–28.

[98] R.-H. Hwang, Y.-L. Hsueh, and H.-W. Chung, "A novel time-obfuscated algorithm for trajectory privacy protection," *IEEE Trans. Services Comput.*, vol. 7, no. 2, pp. 126–139, Apr./Jun. 2014.

[99] S. Gao, J. Ma, W. Shi, G. Zhan, and C. Sun, "TrPF: A trajectory privacy-preserving framework for participatory sensing," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 6, pp. 874–887, Jun. 2013.

[100] G. Theodorakopoulos, R. Shokri, C. Troncoso, J.-P. Hubaux, and J.-Y. Le Boudec, "Prolonging the hide-and-seek game: Optimal trajectory privacy for location-based services," in *Proc. 13th Workshop Privacy Electron. Soc. (WPES)*. New York, NY, USA: ACM, 2014, pp. 73–82.

[101] J. Oksanen, C. Bergman, J. Sainio, and J. Westerholm, "Methods for deriving and calibrating privacy-preserving heat maps from mobile sports tracking application data," *J. Transport Geogr.*, vol. 48, pp. 135–144, Oct. 2015.

[102] Z. Wang *et al.*, "Privacy-preserving crowd-sourced statistical data publishing with an untrusted server," *IEEE Trans. Mobile Comput.*, to be published, doi: 10.1109/TMC.2018.2861765.

[103] Z. Wang *et al.*, "Personalized privacy-preserving task allocation for mobile crowdsensing," *IEEE Trans. Mobile Comput.*, to be published, doi: 10.1109/TMC.2018.2861393.

[104] J. Feng, M. Dijst, B. Wissink, and J. Prillwitz, "Changing travel behaviour in urban China: Evidence from Nanjing 2008–2011," *Transport Policy*, vol. 53, pp. 1–10, Jan. 2017.

[105] X. Kong, M. Li, J. Li, K. Tian, X. Hu, and F. Xia, "CoPFun: An urban co-occurrence pattern mining scheme based on regional function discovery," *World Wide Web*, to be published, doi: 10.1007/s11280-018-0578-x.

**XIANGJIE KONG** (M'13–SM'17) received the B.Sc. and Ph.D. degrees from Zhejiang University, Hangzhou, China. He is currently an Associate Professor with the School of Software, Dalian University of Technology, China. He has published over 70 scientific papers in international journals and conferences (with over 50 indexed by ISI SCIE). His research interests include intelligent transportation systems, social computing, and computational social science. He is a Senior Member of CCF and a member of ACM. He has served as a (guest) editor of several international journals, a workshop chair, or a PC member of a number of conferences.

**MENGLIN LI** received the bachelor's degree in software engineering from the Dalian University of Technology, Dalian, China, in 2016, where she is currently pursuing the master's degree from The Alpha Lab, School of Software. Her research interests include big traffic data mining and analysis, human mobility behavior analysis, and smart city development.

**KAI MA** received the bachelor's degree in software engineering from the Dalian University of Technology, Dalian, China, in 2017, where he is currently pursuing the M.Sc. degree with The Alpha Lab, School of Software. His research interests include vehicular social networks, vehicle data generation, and human behavior.

**KAIQI TIAN** is currently pursuing the bachelor's degree in software engineering with the Dalian University of Technology, Dalian, China. He is also a Research Assistant with The Alpha Lab. His research interests include big traffic data mining and analysis.

**MENGYUAN WANG** received the bachelor's degree in computer science from the Dalian University of Foreign Languages, Dalian, China, in 2017. He is currently pursuing the master's degree from The Alpha Lab, School of Software, Dalian University of Technology. His research interest includes big traffic data mining and analysis.

**ZHAOLONG NING** (M'14) received the M.S. and Ph.D. degrees from Northeastern University, Shenyang, China. He was a Research Fellow with Kyushu University, Japan. He is currently an Associate Professor with the School of Software, Dalian University of Technology, China. He has published over 80 scientific papers in international journals and conferences. His research interests include Internet of vehicles, edge computing, and mobile computing.

**FENG XIA** (M'07–SM'12) received the B.Sc. and Ph.D. degrees from Zhejiang University, Hangzhou, China. He was a Research Fellow with the Queensland University of Technology, Australia. He is currently a Full Professor with the School of Software, Dalian University of Technology, China. He has published two books and over 200 scientific papers in international journals and conferences. His research interests include computational social science, network science, data science, and mobile social networks. He is a Senior Member of ACM and a member of AAAS. He serves as a general chair, a PC chair, a workshop chair, or a publicity chair of a number of conferences. He is the (guest) editor of several international journals.