# Urban Human Mobility: Data-Driven Modeling and Prediction

Jinzhong Wang
[1]Shenyang Sport University
Shenyang 110102, China
[2]School of Software, Dalian University of Technology
Dalian 116620, China
jinz.wang@outlook.com

Xiangjie Kong*, Feng Xia
[2]School of Software, Dalian University of Technology
Dalian 116620, China
{xjkong, f.xia}@acm.org

Lijun Sun
[3]Department of Civil Engineering and Applied Mechanics
McGill University, Montreal, Canada
lijun.sun@mcgill.ca

## ABSTRACT

Human mobility is a multidisciplinary field of physics and computer science and has drawn a lot of attentions in recent years. Some representative models and prediction approaches have been proposed for modeling and predicting human mobility. However, multi-source heterogeneous data from handheld terminals, GPS, and social media, provides a new driving force for exploring urban human mobility patterns from a quantitative and microscopic perspective. The studies of human mobility modeling and prediction play a vital role in a series of applications such as urban planning, epidemic control, location-based services, and intelligent transportation management. In this survey, we review human mobility models based on a human-centric angle in a data-driven context. Specifically, we characterize human mobility patterns from individual, collective, and hybrid levels. Meanwhile, we survey human mobility prediction methods from four aspects and then describe recent development respectively. Finally, we discuss some open issues that provide a helpful reference for researchers' future direction. This review not only lays a solid foundation for beginners who want to acquire a quick understanding of human mobility but also provides helpful information for researchers on how to develop a unified human mobility model.

## 1. INTRODUCTION

Recent years have witnessed explosive growth of people's trajectory data in urban scenarios due to pervasive usage of handheld devices such as smartphones and tablets [132]. Multi-source heterogeneous data acquired by location-acquisition and mobile computing techniques or by the moments that users post on social networking sites, provides an unprecedented opportunity for a deeper understanding of human mobility patterns. Modeling human mobility mainly focuses on exploring the spatiotemporal characteristics and potential regularities hidden in individual and population trajectories. Data-driven modeling becomes increasingly important in a wide range of applications, e.g., epidemic control [9], migratory flows prediction [56, 59, 60], city planning [113], environmental protection [46], and location-based services [117, 118].

Existing research efforts verify that human mobility has 93% potential predictability and random models cannot capture the regularities of human movement accurately [97]. In our daily life, human mobility has a close relationship with users' social contexts as well as urban geography and spatial constraints. For example, mobile users which often co-occur in close proximity may have some correlation between their respective mobility patterns or social properties. Meanwhile, Song *et al.* [96] find that people tend to spend most of their time visiting just a few locations. People usually have short trips for their livelihood rather than long displacements [40]. In addition, human mobility also has other two characteristics, i.e., spatial and temporal. Spatial features reflect a changing law of people geographic location in urban scenarios. Temporal features are dependent on regularities of human mobility in the time dimension. The above-mentioned features contribute to modeling human mobility accurately.

Modeling human mobility has important impacts on different aspects. Human mobility prediction is one of the most important applications in the studies of human mobility. First, it is useful for controlling the spreading of epidemic diseases. These infectious diseases spread through people's travel and interaction. People who are close to the source of infection, are more likely to be infected. Therefore, predicting human mobility can sense the law of migratory flow in advance, and take effective preventive measures, such as crowd evacuation and epidemic control. Second, predicting human mobility is helpful to alleviate traffic congestions and increase travel efficiency. Based on accurate prediction of individual and collective mobility in the future, we can provide travel routes for people to avoid congested roads beforehand. Third, predicting human mobility is very important for commercial applications such as commercial site selection. According to the information of human mobility at different time slots, we can detect potential popular regions and choose the most suitable business location combined with a distribution of POIs (points of interest).

Recently, extensive data-based experiments have been conducted to improve the prediction accuracy of human mobility. Besides, four types of prediction algorithms have been proposed to forecast the next location, next-slot location, and next-slot transition in urban scenarios. Specifically, Markov chain is one of the most popular models, assuming that probability of the next location is dependant on the current and historical sequence of locations. Compression-based methods are proposed to predict human

*Corresponding Author

mobility stemming from a popular incremental parsing algorithm. Time-series methods mainly utilize different technique to improve the performance. Machine learning methods can be used to extract potential regularities by utilizing artificial intelligence and statistical techniques.

In the past few years, several surveys have been presented in the context of human mobility. These surveys review the state of the art at the moment. Aschenbruck *et al.* [6] give a comprehensive survey of publicly available trajectory datasets. The authors discuss some challenges for analyzing trajectory data such as data filtering, data sparsity, validation, and temporal variations. Treurniet *et al.* [110] propose a taxonomy of microscopic mobility models in the field of mobile networking. The taxonomy mainly focuses on the following parts, such as spatial constraints, target selection, pathfinding, motion, pause time, and group dynamics. Hess *et al.* [43] present the specific steps of model building and validation from a perspective of engineering. This paper provides useful information for scholars who want to validate their proposed mobility models. But the above-mentioned contributions are only limited to the domain of mobile communications in mobile social networks and need to be adjusted in urban scenarios. Barbosa *et al.* [10] survey different methods and models that study human mobility patterns from a perspective of machine learning and categorize the existing approaches into three classes. However, the difference from this work is that we not only focus on machine learning but also assess other three analysis methods of human mobility patterns. Toch *et al.* [108] survey the approaches developed to learn various mobility patterns and introduce a range of applications. But related studies on human mobility prediction are not covered. The existing studies have given pointers to different aspects of human mobility, there is no previous thorough survey of challenges and solutions of data-driven modeling and prediction in urban scenarios.

Although an in-depth analysis of multi-source heterogeneous data can provide a comprehensive understanding of human mobility communities, little research has focused on this topic. Prior tools and technologies cannot satisfy the requirement for dealing with multi-source heterogeneous data. Researchers cannot get access to enough travel information. Fortunately, with the development of big data analysis and related technologies, now we can better understand multi-source heterogeneous data and make better use of it. Meanwhile, more and more scholars, institutions, and relative companies have made data available and codes open-source in support replication of experiments.

First, we provide a structured and comprehensive survey of human mobility models and prediction in urban scenarios in this survey as shown in Fig. 1. Second, we divide multi-source heterogeneous data into four major categories as shown in Fig. 2 and analyze the characteristics of various datasets in Table 1. Furthermore, we introduce related work on various data types in detail. Third, we study the properties of human mobility from three perspectives, i.e., spatial scale, temporal scale, and social scale. We also list some evaluation indicators to measure human mobility patterns including trajectory-based, co-occurrence-based, and network-based metrics in Table 2. Then, we study human mobility modeling with an increasing growth of multi-source heterogeneous datasets. We categorize data-driven methods into three classes (individual, population, and hybrid) as

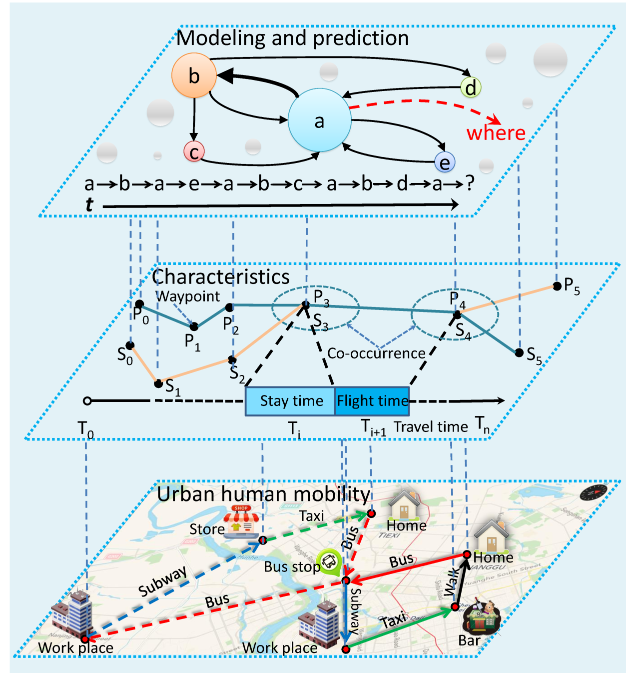


Figure 1: An overview of data-driven modeling and prediction for human mobility.

shown in Fig. 3 and emphasize their principles. Subsequently, we study human mobility prediction methods in urban cities where they aim to improve prediction accuracy. As shown in Fig. 4, we categorize the prediction algorithms into four main classes: Markov-based, Compression-based, Time-series, and Machine learning mechanisms. Finally, we outline their principles, objectives, and datasets of existing literature in Table 3.

Our main contributions can be summarized as follows:

- We present an overview of urban human mobility based on multi-source heterogeneous data and explore social characteristics of human mobility in our daily life.
- We survey recent studies that focus on human mobility models and describe their advantages and disadvantages from the perspective of human mobility characteristics.
- We classify human mobility prediction methods into four categories and introduce their characteristics and research progress.
- We discuss four open problems and emphasize future research trends taking human mobility modeling and prediction into account.

The rest of this paper is organized as follows. In section II, we briefly review a variety of data types for human mobility. Section III mainly presents the three characteristics and fourteen evaluation indicators of human mobility patterns. Section IV provides an overview of data-driven human mobility modeling and introduces their advantages and limitations. Section V introduces representative human mobility prediction methods that aim to improve the prediction accuracy in the field of relevant applications. Section VI

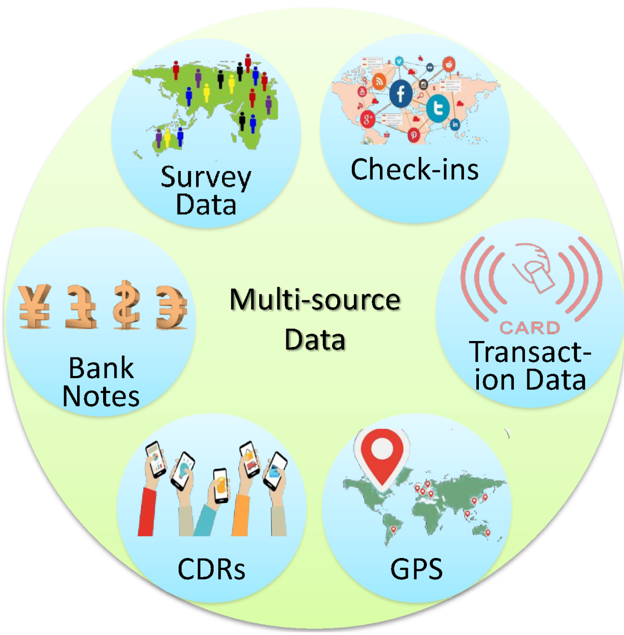


Figure 2: A variety of data types.

discusses several open issues from three aspects. The work is concluded in Section VII.

## 2. MULTI-SOURCE DATA FOR HUMAN MOBILITY

The ubiquitous applications of mobile and handheld devices lead to an explosion of multi-source data correlated with human mobility, providing a novel and comprehensive view to study urban human mobility patterns. These datasets are collected passively, for example, call detail records (CDR), credit card, smart card, their purpose is not collecting mobility data but to register all the transactions. But these multi-source heterogeneous datasets record people's travel trajectories and imply the potential mobility patterns. In smart cities, we can collect a variety of data leveraging network and wireless communication technology, which captures people travel trajectories in their daily life and depicts the spatiotemporal characteristics of urban human mobility. In this section, we introduce the main data types used for human mobility research as shown in Fig. 2 and compare their advantages and disadvantages.

### 2.1 Traditional Research Data

Over the past few decades, due to the singularity of data collection technology, researchers mainly utilize the following data to analyze human mobility patterns, i.e., survey data, bank notes, and CDR. We categorize these 3 kinds of data as traditional data.

#### *2.1.1 Survey Data*

A census is a process of methodically obtaining and recording data about a specific population. National population censuses are used to collect information (gender, age, household, and salary) through answering some questions about the socio-demographic and economics. Particularly, traffic censuses are related to human mobility and acquire some data about daily travel modes, working place, trip duration, and trip displacement. This information contributes to predicting journey-to-work flows and describe commuting behavior.

To the best of our knowledge, there are some surveys to assist with the research of human mobility. In the United States, United States Census conducts a questionnaire survey every ten years since 1790. The data collected contains travel mode and trip time from home to workplace [84]. In the United Kingdom, the national census collects the survey data including names, ages, sexes, occupations, places of birth, locations of workplace and residence for the members of a given population every ten years since 1841. In Australia, travel behavior data is acquired by a statistical census every five years.

In many developed countries, we can download the survey data for free. Specifically, the United States Census Bureau[1] provides the information about the commuting trip between United States' counties. The survey data is very useful for analyzing and modeling human mobility. The data on commuting and migration flows for the United Kingdom is available for free[2]. The data includes more spatial information, i.e., the origin, destination, age, and sex. Meanwhile, we can acquire the survey data for Australia from the website[3]. The data contributes to a deeper analysis from a socio-demographic perspective.

Survey data is invaluable for analyzing human movement and migration. However, the data is only coarse-grained with high costs and statistical errors and is unable to show a dynamic view of human mobility [80].

#### *2.1.2 Bank Notes*

The circulation of bank notes known as bills is closely related to the track of people's activities, which reflects the spatiotemporal characteristics of human mobility to some extent. The website[4] records the trajectories of 464,670 bills in the United States. The data has a total of 250 million records which contain the time and location of bank note transactions. Furthermore, about 11% of the bills are recorded 3 to 5 times. If a bill is registered on the website, its future transactions determine the time and distance between the circulation process. Brockmann *et al.* [18] analyze the spatial characteristics of human travel on geographical scales utilizing the records of circulation. Meanwhile, the results demonstrate that human travel follows a continuous-time random walk model.

The dispersal of bank notes makes a quantitative assessment of human traveling statistics. However, one obvious problem is that the data does not include information about the number of people that have carried a given bill during the two entries in the database. The tracks of bank notes capture the spatiotemporal characteristics of multiple people. Meanwhile, there are sample sparseness and some statistical errors in this data. With this type of data, we can acquire the coarse-grained view of human mobility patterns.

#### *2.1.3 Call Detail Records*

[1]http://www.census.gov/hhes/commuting/data/commuting flows.html
[2]http://webarchive.nationalarchives.gov.uk/20160105160709/
[3]http://www.abs.gov.au/ausstats/abs@.nsf/detailspage/2007.02011?opendocument
[4]http://www.wheresgeorge.com

Call Detail Records (CDR) are records generated by a telephone exchange. The data record includes various information such as time, duration, positions of base stations, and source and destination number [45,82]. When an individual makes a call, the closest cellular network tower that routes the call is recorded, indirectly reflecting the user's geographic location. Due to the widespread use of mobile phones in smart cities, we have access to real-time trajectories of a single individual enabling a deeper study on the spatiotemporal characteristics of human mobility.

In terms of the application of CDR data, Gonzaléz *et al.* [40] study the trace of 100,000 anonymized mobile phone users for a six-month period. To verify their findings, the authors utilize another data set that contains the location of 206 mobile phone users with sampling every two hours for an entire week. The results are that human mobility has a high degree of spatiotemporal regularity, i.e., a time-independent trip displacement and a high probability to return to many frequented locations.

A large-scale CDR data is available for some relevant research in Italy. Except for the basic call information, the data also contains some other information such as weather, news, social media data, and electricity data [11]. The data arises a series of studies on human mobility, which mainly focus on personalized route recommendation [29], capturing individual activity-based behavior [52], and an overview of CDR data analysis [20].

CDR data provides an important source for inferring human mobility but poses some challenges. First, the sampling frequency of CDR data determines the quality of experimental results. There is the inherent anisotropy of call patterns. Many calls are made in a short time, but there are no activities in the following periods. Therefore, it is the issue on how to select an appropriate time bin. Second, the accuracy of user's location needs to be further improved. In urban cities, the coverage areas of cell network towers vary from tens of meters to several kilometers. The position of cell towers approximately represents the geographic location of the user making a call. In general, the positioning error ranges from 100 to 1000 meters.

## 2.2 Popular Urban Data

As opposed to traditional data, popular urban data refers to human trajectory datasets collected by mobile communication and wireless network technology. These types of data provide a fine-grained view of human mobility patterns and contribute to a deeper understanding of human travel characteristics.

### *2.2.1 GPS Data*

Global Positioning System (GPS) provides geolocation information to a GPS user with at least four GPS satellites. In smart cities, smartphones and vehicles equipped with a GPS receiver can record the trajectories of user's movement with a high degree of accuracy and continuous spatiotemporal resolution. Therefore, GPS data is able to provide an important source which is used to explore human mobility patterns.

GPS data arouses widespread concerns in the study of human travel behavior. Rhee *et al.* [90] find statistically similar features between human walk patterns and Levy walks based on 226 GPS trajectories. Specifically, the authors choose five sites for acquiring human mobility traces. The data records the location information every 10 s with a position accuracy of 3 m. In addition, the taxi GPS data is also available for the study in Beijing, China[5]. The data collects 12,000 taxis running trajectories during November 2012, including time, latitude, longitude, and service status. The relevant studies include traffic congestion estimation and prediction [60], and taxi service recommendation [59]. Meanwhile, Kong *et al.* [57] detect long-term traffic anomalies by using the bus GPS data in Hangzhou, containing temporal and spatial information of bus trajectories. Alessandretti *et al.* [4] consider the preferential and explorative characteristics of human behavior and find the correlation between the size of individual preferred locations and their number of social interactions. Moreover, Microsoft releases the GPS data generated from 182 individuals' daily trajectories[6]. The data contains 17,621 traces with a high spatiotemporal resolution recorded every 1-5 s. Based on the data, the studies include exploring human mobility [133], mining interesting locations [135], exploiting user similarity [65], and social networking service [134].

In terms of the positioning accuracy, the error of GPS data is much lower than CDR data, which provides more accurate location information for GPS enabled mobile devices. However, GPS transmitters do not work without the supply of power, which may lead to the loss of people's continuous spatiotemporal mobile location information.

### *2.2.2 Public Transport Transaction Data*

In urban cities, public transportation system is becoming more and more developed and makes people's lives more convenient. People usually carry a smart card to travel for social activities, which generate massive trip information, e.g. card ID, trip origin, boarding time, trip destination, alighting time, and trip expense. The data captures people's travel behavior precisely and provides a new source for exploring human mobility patterns.

Transport transaction data has received a lot of attention. Le *et al.* [62] use smart card data to identify different classes of transit passenger from a perspective of behaviors and needs. Hong *et al.* [44] provide a new method to detect urban black holes and volcanos based on the bike trip data in Manhattan. Du *et al.* [31] design a suspect detection and surveillance system to identify pickpocket based on their daily transaction data in transit systems. Zhang *et al.* [129] develop a novel method to extract spatiotemporal segmentation information such as transfer time. Sun *et al.* [103] propose a probabilistic tensor factorization framework to understand urban human mobility using on 14 million transit records. Zhao *et al.* [130] focus on the pattern of passenger route choice and propose a probabilistic model to estimate the displacement of passengers in Shenzhen subway system. Mahrsi *et al.* [68] propose two approaches to cluster smart card data, which mainly consider passengers' activity and boarding times. Wang *et al.* [113] analyze Shanghai subway transaction data and identify 10 functional clusters of subway stations. Xia *et al.* [121] leverage multi-source heterogeneous transport data to explore the patterns of human mobility in Shanghai, China. Sun *et al.* [104] study people's

[5]http://www.datatang.com/data/44502
[6]http://research.micosoft.com/enus/downloads/b16d359d-d164-469e-8fd4-daa38f2b2e13

encounter mechanisms and structures from a perspective of collective and individual scales based on public bus datasets in a city.

Transport transaction data is collected by automated fare collection systems and has a high accuracy of data. However, the data only covers the span of traveling by public transport, which leads to an incomplete view of human mobility. Meanwhile, we cannot acquire information on the passenger's gender, age, and work unit considering the privacy of cardholders. To explore human mobility patterns more deeply, we need to combine transport transaction data with other types of data.

#### 2.2.3 Social Media Check-in Data

Social media check-ins data is also a valuable source collected by social network providers. More and more users record their social connections and geographical locations through Twitter, Facebook, Foursquare, and Flickr. The geotagged data is made up of coordinates, time, photos, and comments. Therefore, the movement trajectories of users can be obtained from the sequence of published locations. Through analyzing the information, we can calculate some important metrics, i.e., radius of gyration, jump length, visit frequency. Scellato *et al.* [93] investigate three location-based social networks from a perspective of network science and discover several small-world properties. Itoh *et al.* [48] explore social media data on Twitter and smart card data on the Tokyo subway and extract abnormal situations on mega-city subway networks. Ni *et al.* [74] propose a hybrid approach to predict the subway passenger flow based on social media activities and improve the prediction accuracy.

Compared with other data sources, social media data has its unique characteristics such as more social information, which provides a multidimensional view of studying human mobility patterns. But the data is uploaded personally, so it is likely to have false information.

To illustrate various data types more clearly, we summarize their advantages and disadvantages as shown in Table 1.

## 3. HUMAN MOBILITY CHARACTERISTICS AND METRICS

In this section, we focus on the characteristics of human mobility patterns from three aspects, i.e., temporal scale, spatial scale, and social scale. Meanwhile, we illustrate several important metrics to measure the patterns of human mobility referred to Table 2.

### 3.1 Characteristics

Human mobility refers to the movement of people in a spatio-temporal dimension. Meanwhile, people may meet or socialize at a certain time and location during the process of moving. So we describe the characteristics of human mobility from the following three aspects.

#### 3.1.1 Temporal Scale

Temporal scale is one of the characteristics of human mobility patterns. In urban cities, people usually travel for your social activities such as commuting, shopping, traveling, and dating. The daily activities reflect the statistical regularities of time. For example, someone often goes to work by bus at 7:00 on weekdays. Then we will have relevant questions on the laws of time. How long does this trip take? Which distribution does the trip interval follow? How often does this person ride a bus? To answer these questions, we need to analyze the temporal scale prosperities hidden in various data.

To the best of our knowledge, there is a lot of literature focusing on the temporal scale of human mobility [3, 23, 111, 114, 121]. Xia *et al.* [121] conclude that the law of trip duration is fitted to log-normal distribution for taking a taxi and Weibull distribution for taking a subway based on the empirical analysis. Csáji *et al.* [23] discover log-normal distribution fitting trip duration and trip interval based on taxi trajectory data. Veloso *et al.* [111] propose exponential distribution fitting the interval of two consecutive trips. Wang *et al.* [114] infer that log-normal distribution fits trip duration and trip interval based on taxi trajectory data from five cities. Alessandretti *et al.* [3] verify that waiting time between consecutive locations are best fitted by gamma distributions.

#### 3.1.2 Spatial scale

Spatial scale is closely related to the act of human displacement. Human mobility mainly includes short distance trip (intra-city) and long distance migration (inter-country). An individual moves from an origin to a destination and forms a travel trajectory on a spatial scale. In smart cities, collective spatial trajectories provide a huge amount of spatial information which contribute to further analyzing human mobility patterns. Currently, scholars concentrate on some spatial metrics, e.g., trip displacement, radius of gyration, and spatial entropy.

Spatial regularity has received a considerable amount of attention. Pearson *et al.* [81] first propose a term "random walk" which illustrates a path with a succession of random steps and lay a solid foundation for many scientific fields. Meanwhile, lévy flight model infers the distance following the heavy-tailed distribution [40, 90]. Brockmann *et al.* [18] propose a power law distribution fitting to human travel distance based on the circulation of bank notes in the United States. Song *et al.* [97] conclude that 93% of human mobility patterns have the characteristics of predictability by analyzing mobile phone user's entropy. Zhao *et al.* [131] discover log-normal distribution for trip distance by single transit mode and power law distribution for the mixed mode.

Calabrese *et al.* [19] discover power law with exponential cut-off fitting to trip displacement based on traffic trajectory data. However, Wang *et al.* [114] propose trip displacement follows exponential distribution rather than power distribution by analyzing taxi trajectory data from five cities. Veloso *et al.* [111] verify gamma distribution fitting to trip distance by analyzing a taxi dataset in Lisbon. Csáji *et al.* [23] study the commuting distances fitted to log-normal distribution using a mobile phone dataset in Portugal.

#### 3.1.3 Social Scale

Social scale is another feature related to human mobility because human beings are social animals. In our daily life, there is a phenomenon called as co-occurrence that people from different districts often meet or locate within a specified distance at the same time slot [83]. In other words, we can entitle them as "familiar strangers". The patterns are valuable for a range of applications including social network [55], routing [117, 118], spread of information [49, 119], and innovations [115].

Table 1: Comparisons of a variety of data

| | Data type | Advantages | Disadvantages |
|---|---|---|---|
| **Traditional** | Survey data | Free; Socio-demographic information | High costs; Statistical errors; Static |
| | Bank notes | Trajectory information | Sample sparseness; Statistical errors; Coarse-grained |
| | Call detail records | Trajectory information | Positioning error |
| **Popular** | GPS data | Fine-grained; High positioning accuracy | High battery usage |
| | Public transaction data | High positioning accuracy | Coarse-grained |
| | Social media check-in data | More contextual information | Human errors |

Crandall *et al.* [22] discover a positive correlation between the probability of a social tie and the number of co-occurrences among Flickr users. Hsieh *et al.* [47] leverage a co-location graph to construct users' indirect linkages and identify paired friendship. Deville *et al.* [28] discover a scaling law measuring human movement and spatial communication and decrease the number of independent parameters. Yu *et al.* [126] propose a novel method to forecast people's occasional social interaction based on GPS trajectory data. Xia *et al.* [117] design a social-based routing protocol (PIS) to search intermediate transmitter and improve the performance of PIS. Yi *et al.* [124] propose a mobility based model to distinguish co-occurrences between acquaintances and strangers.

## 3.2 Metrics

In this section, we illustrate important evaluation metrics characterizing mobility patterns from the following three aspects.

### 3.2.1 Trajectory-based

Trajectory is the most intuitive expression of human mobility and contains many features and laws of human behaviors.

- Displacement: Displacement is the most fundamental quantity to measure the distance of a path moved by an individual. We can obtain the length by computing the Euclidean distance for origins and destinations' coordinates or the geodesic distance for its longitude and latitude [121]. Displacement includes short-term path and long-term path, characterizing human movement dynamics. Meanwhile, the metric is collected by the circulation of bank notes, call detail records, and social check-ins data. Some scholars propose that displacement is followed by a power law distribution by conducting extensive experiments [18, 131].

- Duration: Duration is the elapsed time during a journey, i.e. the difference between the departure time and the end time. The metric reflects the time spent in visiting locations, capturing the basic characteristics of human mobility. We also leverage the metric to determine an individual's home and workplace [94].

- Speed: Speed is the metric of people's travel efficiency, measuring the relationship between the above-mentioned quantities. In urban cities, travel speed usually fluctuates because traffic conditions are complex [121]. For example, time spent on a given trip is not proportional to the distance traveled in the rush hours. People have to go out ahead of time or take into account the mode of transportation in the case of traffic congestion.

- Interval: Interval is the elapsed time between two consecutive trips, which is different from duration [114]. For taxi operation, taxi drivers cruise for a new passenger after passengers get off. So the metric measures the time of non-occupied state for taxi and reflects travel demands indirectly, i.e. the degree of taking a taxi. Obviously, a taxi with shorter interval will make more money.

- Radius of gyration: Radius of gyration is the root mean square distance of an individual's location from his center of mass of the motions [40]. It is defined as Equation 1.

$$Radius_g = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(r_i - r_o)^2}, \tag{1}$$

where $r_i$ represents the $i$th position of a user, and $r_o$ is the center of mass of his/her travel trajectories, $r_o = \sum_{i=1}^{n} r_i/n$. Radius of gyration is an important metric to measure the typical distance traveled by people and it has a close relationship to the mutual distance of visited locations and the total number of visits [96]. Radius of gyration provides a more complete view of how individuals travel around their centers of mass.

- Entropy: Entropy is the most basic metric measuring the degree of predictability of an individual's human mobility [97]. It is related to the frequency of visitation and capture the full spatiotemporal pattern in a person's travel trajectory. People with a high entropy show a low heterogeneity of visiting location. We formulate the metric as follows:

$$Entropy_g = -\sum_{i=1}^{n} p_g(i)\log_2 p_g(i), \tag{2}$$

where $p_g(i)$ denotes the historical probability that the user $g$ visits the location $i$, characterizing the heterogeneity of visiting patterns. $n$ represents the number of visited locations.

Table 2: Summary of the characteristics and metrics of human mobility patterns

| Characteristics | Reference | Trajectory-based | | | | | | Co-occurrence-based | | | | Network-based | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **Trip displacement** | **Trip duration** | **Trip speed** | **Trip interval** | **Radius of gyration** | **Trip entropy** | **Frequency of co-location** | **Closeness of significant locations** | **Probability of co-location** | **Similarity of location history** | **Degree centrality** | **Betweenness centrality** | **Closeness centrality** | **Eigenvector centrality** |
| Spatiotemporal scale | Xia *et al.* [121] | • | • | • | • | | | | | | | | | | |
| | Veloso *et al.* [111] | • | • | | | | | | | | | | | | |
| | Wang *et al.* [114] | • | • | | • | | | | | | | | | | |
| | Alessandretti *et al.* [3] | • | | | • | | | | | | | | | | |
| | Pearson *et al.* [81] | • | | | | | | | | | | | | | |
| | Gonzalez *et al.* [40] | • | | | | • | | | | | | | | | |
| | Rhee *et al.* [90] | • | • | | • | | | | | | | | | | |
| | Brockmann *et al.* [18] | • | • | | | | | | | | | | | | |
| | Song *et al.* [97] | | • | | • | | • | | | | | | | | |
| | Zhao *et al.* [131] | • | • | | | | | | | | | | | | |
| | Calabrese *et al.* [19] | • | | | | | | | | | | | | | |
| | Song *et al.* [96] | | | | | • | | | | | | | | | |
| | Csáji *et al.* [23] | • | | | | • | | | | | | | | | |
| Social scale | Crandall *et al.* [22] | | | | | | | • | | • | | | | | |
| | Hsieh *et al.* [47] | | | | | | | • | • | • | • | | | | |
| | Deville *et al.* [28] | • | | | | | | | • | | • | | | | |
| | Yu *et al.* [126] | | • | • | | | | | | • | | | | | |
| | Xia *et al.* [117] | | | | | | | | • | | • | | | | |
| | Yi *et al.* [124] | | | | | | | • | • | | | | | | |
| | Newman *et al.* [73] | | | | | | | | | | | • | • | • | • |
| | Rahim *et al.* [88] | | | | | | | | | | | • | | | |
| | Freeman *et al.* [37] | | | | | | | | | | | | • | | |
| | Daly *et al.* [25] | | | | | | | • | • | | • | | • | | |
| | Freeman *et al.* [38] | | | | | | | | | | | • | • | • | |
| | Bonacich *et al.* [16] | | | | | | | | | | | | | | • |
| | Blanford *et al.* [15] | • | | | | • | | | | | | | • | • | • |

### 3.2.2 Network-based

The multi-source heterogenous data provides a novel view of studying human mobility from a perspective of complex network [40, 97]. We can model human behavior as a graph to explore the characteristics of human dynamics. In the graph, the nodes represent a set of locations or POIs visited by people, and an edge represents the related pairs of locations in historical trajectories. Furthermore, it can be determined which node is the most influential in a network and how well a network is optimized with respect to network performance.

- Degree Centrality: Degree centrality is defined as the number of ties that a node has [73]. Degree centrality is shown in Equation 3.

$$C_D(i) = \sum_{j=1}^{n} a_{ij}, \tag{3}$$

where $C_D(i)$ denotes the degree centrality of node $i$. $\sum_{j=1}^{n} a_{ij}$ is the number of links between node $i$ and the other nodes (n-1). Degree centrality identifies the most important nodes in a network. A node with high degree centrality has more links than others, which is used for modeling human dynamics [88].

- Betweenness centrality: Betweenness centrality measures the number of times a node acts as a bridge along the shortest path between two other nodes [37]. The betweenness centrality is computed as follows:

$$C_B(k) = \sum_{i \neq k \neq j \in V} \frac{\alpha_{ij}(k)}{\alpha_{ij}}, \tag{4}$$

where $C_B(k)$ represents the betweenness centrality of node $k$. $\alpha_{ij}$ is the total number of shortest path from node $i$ to node $j$ and $\alpha_{ij}(k)$ is the number of these paths through node $k$. According to its definition, we can discover that a node with high betweenness centrality greatly impacts the transmission of information flowing between others [25]. Therefore, it is important to prevent nodes with high betweenness centrality from failing in a network.

- Closeness centrality: Closeness centrality is defined as the reciprocal of the sum of the shortest path between a given node and all other nodes in the network [38]. We formulate closeness centrality as follows:

$$C_C(i) = \frac{1}{\sum_{j \neq i} d(j,i)}, \tag{5}$$

where $C_C(i)$ is the closeness centrality of node $i$. $d(j,i)$ is the sum of between node $j$ and node $i$. A node with a high closeness centrality is clearer to all the other nodes in the network, and its movement law reacts to the motion law of other nodes. Therefore, the patterns of high closeness centrality node are critical to predicting human mobility in the future.

- Eigenvector centrality: Eigenvector centrality is a measure of the influence of a node in a network. It calculates relative values to all nodes in a network [16]. The underlying assumption is that high-score nodes are more useful to improve their neighbor nodes' scores than low-scoring nodes.

$$C_E(i) = \lambda \sum_{j=1}^{n} a_{ij} x_j, \tag{6}$$

where $\lambda$ is a constant, and $\langle a_{ij} \rangle$ is the adjacency matrix of the network. for studying movement patterns and mapping inter/intra-region movement dynamics. It can measure the importance of a node based on the node's connections in a network. In the field of human mobility, the metric is used for illustrating the degree of influence of each district [15].

### 3.2.3 Social-based

Location information data implicitly characterize people's social behavior from a perspective of social networks. The probability of social interaction contributes to improve the accuracy of human mobility prediction [47].

- Frequency of co-occurrence: It is the number of their co-occurrence at a specified time slot. Two people with higher meeting frequency tend to be friends. The metric is defined as follows:

$$F_{i,j} = |N_{i,j}|, \tag{7}$$

where $i$ denotes an individual. $N$ is the number of their co-locations.

- Closeness of important locations: If the locations visited by two people are closer to one another, they have a high likelihood of having the friendship. We leverage $LI(i,l)$ as the location importance of location $l$ for an individual $i$. In addition, We use $l^*(i) = \arg\max_{l \in L_i} LI(i,l)$ as the most important location of $i$. Closeness of important locations is defined as:

$$D_{i,j} = dist(l^*(i), l^*(j)), \tag{8}$$

where $dist(,)$ denotes the geodesic distance between the most important locations visited by two people.

- Probability of co-occurrence: The probability of co-occurrence measures the chance of two people visiting the same location in the future. We define the metric as Equation 9.

$$P_{i,j} = \sum_{l \in L_i \bigcap L_i} LI(i,l) \times LI(j,l). \tag{9}$$

- Similarity of historical trajectories The metric measures the similarity degree of historical trajectories of two people. It is defined as:

$$S_{i,j} = \sum_{l \in L_i \bigcap L_j} \frac{LI(i,l) \times LI(j,l)}{\|LI(i,l)\| \times \|LI(j,l)\|}. \tag{10}$$

If the historical trajectories of two people are much alike to each other, they tend to be acquaintances with respect to these locations and its importance.

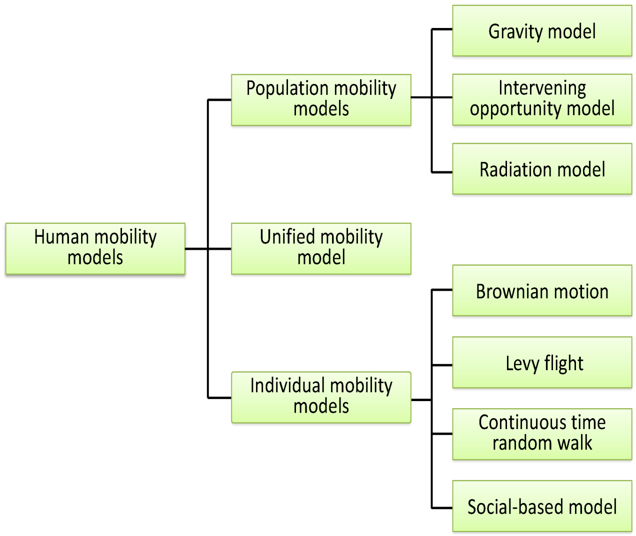


Figure 3: Taxonomy of data-driven human mobility models.

# 4. DATA-DRIVEN MODELING

In recent years, the availability of various types of data characterizing aspects of human behavior has provided a great opportunity to modeling human mobility. In this section, we summarize the development of the subject and review the state-of-the-art mobility models as shown in Fig. 3, e.g. population mobility model, individual mobility model, and unified mobility model.

## 4.1 Population Mobility Model

Population mobility model mainly focuses on the mobility patterns of collective people between two regions in urban scenarios. This type of model can predict the distribution of migratory flow at some time in the future according to the population of regions.

### *4.1.1 Gravity model*

The gravity model is inspired by Newton's law of gravitation [136]. It assumes that the volume $T_{ij}$ of people flow between locations $i$ and $j$ is in direct proportion to the population size of the two locations and is in inverse proportion to the distance $d_{ij}$ between them [5,12]. The gravity model is defined as

$$T_{i,j} = \frac{x_i^{\alpha} x_j^{\beta}}{f(d_{ij})}, \tag{11}$$

where $\alpha$ and $\beta$ are two exponents. $x_i$ and $x_j$ denote the population of locations $i$ and $j$. $f(d_{ij})$ is a function of distance between origins and destinations, which can approximate the empirical data such as power law and exponential.

In addition to estimating human flow, the gravity model is used for measuring the intensity of interaction (calls and trade) between two regions. The form of the function is tunable according to the scenarios. In urban planning, the distance may not measure the travel cost between specified locations, and trip duration may be a better alternative. The gravity model is widely used in public transportation management [54], geography [36], telecommunication [61], and social economics [70].

Although the gravity model is very popular, it has several limitations. The gravity model is a coarse-grained characterization of human mobility. The model needs to estimate some parameters and functions based on the empirical data in advance [64, 94]. Meanwhile, the model assumes that the flux between two locations is the same for both sources and destinations, which is obviously inconsistent with reality. The model is not sensitive and is unable to give reasons for fluctuations in the number of travelers between two locations.

### *4.1.2 Intervening opportunity model*

The intervening opportunity model is proposed in 1940 by Stouffer [102]. The model assumes that the number of individuals $T_{ij}$ that move between locations $i$ and $j$ is proportional to the number of opportunities at that distance, and decays with the number of intervening opportunities. The probability of migration is impacted by the opportunities to a destination rather than distance and population size. The model is defined as

$$T_{i,j} = p_i \frac{e^{-\alpha S_{ij-1}} - e^{-\alpha S_{ij}}}{1 - e^{-\alpha S}}, \tag{12}$$

where $p_i$ is the number of people originated from location $i$. $S_{ij}$ is the number of people between locations i and j and including the two locations ranked by travel cost from the origin. $S$ is the total number of people. $\alpha$ denotes the probability of accepting an opportunity destination.

The intervening opportunity model is a conceptual model to explore human mobility patterns and receives a lot of attention [21, 91, 116]. Furthermore, the model reflects the decision process of the selection of destinations. However, unlike the gravity model, the intervening opportunity model is more complex and tends to underestimate the proportion of long distance [77].

### *4.1.3 Radiation model*

The radiation model is put forward by Simini [94]. The model compensates for the flaw of the gravity model that cannot explain the difference in the bi-directional migration flows between origins and destinations. Furthermore, the radiation model only requires geographical information on population size without a pre-estimation of parameters [95]. It is defined as

$$T_{i,j} = p_i \frac{x_i x_j}{(x_i + N_{ij})(x_i + x_j + N_{ij})}, \tag{13}$$

where $p_i$, $x_i$, and $x_j$ are the same to the representation of above-mentioned models. $N_{ij}$ represents the total number of population between locations $i$ and $j$ (other than $i$ and $j$). The specific range is a circle with location $i$ as its center and $d_{ij}$ as its radius, excepting for locations $i$ and $j$.

The radiation model has a good performance on predicting long-term migration rather than short-range travel mobility patterns [66, 79]. However, the information of population size is not easily obtained due to human mobility dynamics. In urban cities, people usually travel for their social activities such as working, shopping, and traveling. Therefore, it is not appropriate to utilize population distribution to characterize individual mobility patterns [64, 69, 89].

## 4.2 Individual Mobility Model

Individual mobility model mainly characterizes the mobility patterns of individuals based on multi-source spatiotemporal trajectory datasets. This type of model introduces some travel features to modeling human mobility, i.e., radius of gyration, entropy, trip interval, and trip displacement.

#### 4.2.1 Brownian motion

Brownian motion originally explains the motion law of particles hovered in a liquid [30, 35]. The process of motion is accompanied by multiple collisions and stochastic steps and directions. The probability density function is represented as

$$P(x,t) = \frac{1}{\sqrt{2\pi t\sigma^2}} e^{\frac{-(x-\mu kt)^2}{2t\sigma^2}}, \tag{14}$$

where $\sigma^2$ and $\mu$ denote the variance and mean of the stochastic distances.

Brownian motion is also a kind of random walk. We can define Brownian motion as a limit of the non-continuous symmetrical random walk. Einstein *et al.* [32] illustrate the principle of Brownian motion in detail and provide a strong evidence of the existence of atoms and molecules.

#### 4.2.2 Lévy flight

Lévy flight is proposed by analyzing the traces of bank notes [18]. The model discovers that human mobility consists of a number of short displacements, mixed irregularly by a long displacement. The flight step $l$ is approximated by a power law with the characteristics of a fat-tailed distribution.

$$P(l) \sim l^{(1+\beta)}, 0 < \beta < 2 \tag{15}$$

where $\beta$ is an exponent of displacement.

To explain the intrinsic mechanisms of Lévy flight, some literature makes a series of extensive experiments. Gonzales *et al.* [40] prove that the power-law distribution of jump length is related to the characteristics of individual mobility and population heterogeneity. Han and Jiang *et al.* [42] demonstrate that human mobility is impacted by the topology of urban transportation systems. Jiang *et al.* [51] also find the above-mentioned factor by analyzing the human mobility in urban road networks.

#### 4.2.3 Continuous time random walk

The two models mentioned above are a class of random walk with a sparse time dimension. A jump step is determined by the length distribution of human mobility. However, a continuous time random walk (CTRW) shows the characteristics of random jump lengths and time intervals in a time slot [18].

The CTRW is composed of a spatial displacement $\Delta x$ and a temporal increment $\Delta t$. After $n$ iterations the position of an individual is calculated by

$$X_n = \sum_{i=1}^{n} \Delta x_i, \tag{16}$$

and the time interval corresponds to

$$T_n = \sum_{i=1}^{n} \Delta t_i. \tag{17}$$

We utilize the Fourier-Laplace transform to represent the process of CTRW [72].

$$W(k,u) = \frac{1-\phi(u)}{u(1-\phi(u)f(k))}, \tag{18}$$

where $\phi(u)$ and $f(k)$ is the Laplace and Fourier transform of the probability density function of $\Delta t$ and $\Delta x$, respectively.

#### 4.2.4 Social-based models

The ubiquity of mobile devices and the rise of location-based services provide a novel perspective for modeling human mobility patterns. Social-based models mainly utilize people's locations and social interaction information to characterize individuals' behaviors.

Crandall *et al.* [22] propose a spatiotemporal co-occurrence model verifying the relationship between the frequency of co-occurrence and the presence of social ties. The model is defined as

$$P(F|C_d) = \frac{1}{M} e^{d \log \beta(N-1)+1}, \tag{19}$$

where $N$ is the number of divided geographic grids. $M$ is the number of population. $\beta$ is the probability of visiting a place by pairwise friends. $P(F|C_d)$ denotes the probability that two individuals have a social relationship (friends, relatives, and colleagues) given that they visit the same place on $d$ days. Meanwhile, the growth rate of the exponential function is dependant on both $N$ and $\beta$.

Pham *et al.* [83] propose an entropy-based model (EBM) which demonstrates their social connections and social tie strength by analyzing people's co-occurrences in urban cities. The EBM model is represented as

$$s_{i,j} = \alpha[\sum_{l,c_{ij},l\neq 0} (\frac{c_{ij,l}}{f_{ij}})^q]^{1/(1-q)} + \beta \sum_l c_{ij} \times \exp(-H_l) + \gamma, \tag{20}$$

where $\alpha$, $\beta$, and $\gamma$ are parameters determined by experimental datasets. $c_{ij,l}$ is the set of co-occurrences of user $i$ and $j$ at location $l$. $q$ is the order of diversity showing its sensitivity to $c_{ij,l}$. $H_l$ is the entropy of location $l$.

Yan *et al.* [47] design a two-stage feature engineering framework to discover social ties based on people's geographical footprints. The researchers make a range of experiments on Gowalla check-in data and verify the effectiveness of their framework. Deville *et al.* [28] analyze three mobile phone datasets and find a new scaling relationship between human mobility and social networks.

Noulas *et al.* [76] explore urban human mobility patterns based on Foursquare datasets and develop a rank-based movement model to accurately characterize real human movements in several cities. The model of human mobility is defined as

$$P_r[u \to v] \propto \frac{1}{|w : d(u,w) < d(u,v)|^{\alpha}}, \tag{21}$$

where $P_r[u \to v]$ is the probability of moving from location $u$ to location $v$. $d(u,w)$ is the distance between location $u$ and location $w$. $|\,|$ is the number of places meeting a condition. $\alpha$ is an exponent to scale human mobility patterns for different cities.

### 4.3 Unified Mobility Model

Although there is a large amount of above-mentioned research, modeling human mobility patterns from both the individual and population levels is still an open issue. Yan

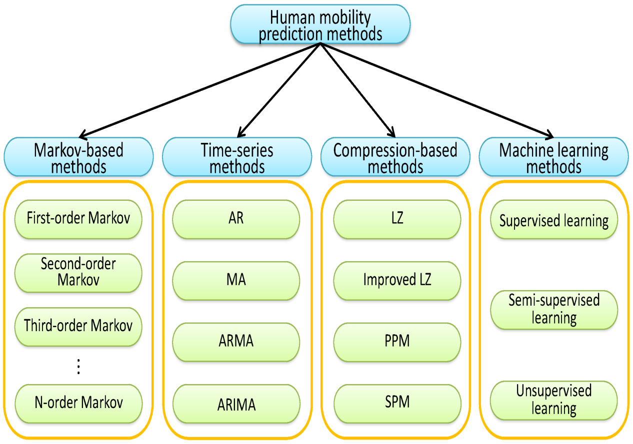


Figure 4: Taxonomy of human mobility prediction algorithms.

*et al.* [122] propose a universal model by integrating memory effect and population-induced competition to improve the prediction accuracy of human mobility. The probability $p_{ij}$ of moving from location i to j is defined as

$$P_{ij} \propto \frac{m_j}{W_{ji}}(1 + \frac{\lambda}{r_j}), \tag{22}$$

where $\lambda$ is a parameter that measures the strength of the memory effect. $r_j$ means that location $j$ is the $r$th visited location in people's travel trajectories, showing the original attractiveness of the location. $W_{ji}$ is the total population in a circle centered at $j$ with a radius of the distance between locations $j$ and $i$. $m_j$ is the population of location $j$.

For the model only requires the distribution of population, it can provide an alternative approach to extracting the important mobility patterns with reasonable accuracy. Furthermore, it can be used in a range of application such as relieving traffic congestions, predicting migration influx, planning urban facilities, and understanding other human dynamic behaviors.

# 5. HUMAN MOBILITY PREDICTION ALGORITHMS

Human mobility prediction is a crucial problem in smart cities and has a close relationship with a large number of practical applications such as intelligent transportation management [57, 59, 60, 120], city planning [113, 120, 121], epidemic control [8, 33, 107], emergency management [50, 67, 92], disaster management [101], and social networking services [58, 75, 115]. In this section, we divide prediction algorithms into four categories and illustrate them in the following subsections as shown in Fig. 4. Meanwhile, we summarize the relevant research on human mobility prediction methods in Table 3.

## 5.1 Markov-based methods

Markov-based methods are a representative of stochastic models that illustrate the probabilities of transition from one state to another. For the observations, Markov-based methods assume each observation to be a state. If the transition is dependant on the present state, the model is called as the 1st order Markov model. The 2nd order Markov model considers both the current and previous states given a sequence of observations.

Song *et al.* [98] focus on the next-location prediction of human mobility and verify the superiority of Markov methods compared with other prediction methods (compression-based, partial matching, and sampled pattern matching). Qiao *et al.* [86] propose a hidden Markov model which has the characteristics of self-adaptive parameter selection according to the change of objects' moving speed. The experimental results show a high positioning precision subjected to stochastically changing speeds. Terroso *et al.* [105] develop a multilevel Markov-based approach to predict the next position of people and verify its effectiveness by using geo-tagged tweets. Qiao *et al.* [87] design a hybrid Markov model considering the spatiotemporal characteristics of human mobility and predict people's future locations with about 56% accuracy. Yin *et al.* [125] propose a hidden Markov models to extract travelers' activity patterns from CDRs and provide an actionable and modular solution to analyze people's travel demand.

## 5.2 Compression-based methods

Compression-based methods include Lempel-Ziv (LZ), improved LZ, Partial Matching (PPM), and Sample Pattern Matching (SPM), which are originally used in text compression. The type of methods stems from a popular incremental parsing algorithm proposed by Ziv *et al.* [137]. Furthermore, compression-based methods are like the $k$th Markov model except that $k$ can be infinite [14].

Some researchers leverage the above-mentioned methods to improve the prediction accuracy of human mobility. Alam *et al.* [2] propose an enhanced episode discovery prediction algorithm to predict inhabitant activities and achieve 88.3% prediction accuracy. Gopalratnam *et al.* [41] propose an active LeZi prediction algorithm to predict the future events by using a sequence of previous events. Pulliyakode *et al.* [85] propose an improved PPM algorithm to predict people future location and demonstrate the effectiveness based on a location dataset.

## 5.3 Time-series methods

Time-series methods are applied to time series data to predict future values in the series. Given a time series of data, these algorithms leverage different techniques to make times series data stationary, which contributes to the improvement of prediction accuracy. These methods include Autogressive (AR) [112, 127], Moving Average (MA) [109], Autoregressive Moving Average (ARMA) [26], Autoregressive Integrated Moving Average (ARIMA) [17].

Yang *et al.* [123] propose a novel method (time series models) to predict future links in human mobility networks and acquire a greater precision accuracy. De *et al.* [27] utilize multivariate nonlinear time series methods to predict human movement by combining with human mobility patterns. Zeng *et al.* [128] propose a high-order time series model to predict the visited locations in the future, and also study the predictability by analyzing other factors such as spatiotemporal resolution, radius of gyration.

## 5.4 Machine learning methods

Machine learning methods are widely used for human mobility prediction. Machine learning methods mainly utilize

Table 3: Summary of the methods of human mobility prediction

| Taxonomy | Reference | Principle of proposed algorithms | Objective |
|---|---|---|---|
| Markov-based | Song *et al.* [98] | The $O(2)$ Markov predictor assumes that the next location can be predicted from the last two contexts | Next location |
| | Qiao *et al.* [86] | A hidden Markov model to predict the position and behaviorof moving objects | Continuous traces |
| | Terroso *et al.* [105] | A multilevel Markov-based method that utilizes mobile andsocial media datasets | Next activity and position |
| | Qiao *et al.* [87] | A hybrid Markov-based model that considers the spatiotemp-oral and similarity of human mobility patterns | Next location |
| Compression-based | Alam *et al.* [2] | A partial matching algorithm to predict the next activity from the previous history | Next activity |
| | Gopalratnam *et al.* [41] | An active LeZi algorithm that utilizes an observed sequence of events to predict the next event | Next event |
| | Pulliyakode *et al.* [85] | An improved PPM algorithm that assumes frequency trees of greater depth resulting in better prediction | Next location |
| Time-series | Yang *et al.* [123] | A link prediction technologies that uses network density timeseries and edge life to infer future links | Next link |
| | De *et al.* [27] | A multivariate nonlinear time series prediction techniques bytaking movements of friends and people into account | Social Interactions |
| | Zeng *et al.* [128] | A time-series-based method using Gibbs sampling to predict users's movement | Next location |
| Machine learning | Song *et al.* [99] | An intelligent system (DeepTransport) based on the deep learning architecture to predict human mobility in urban scenarios | Future movements transportation mode |
| | Feng *et al.* [34] | An attentional recurrent neural network that captures heterogeneous transition regularity and multi-level periodicity | Next location |
| | Baumann *et al.* [13] | A generic framework that explores human mobility data andcompute both population models and individual models to predict human mobility | Next place<br>Next-slot place<br>Next-slot transition |

statistical techniques to extract potential laws in specific data. By mining people's trajectory data, machine learning methods can predict the volume of migration flow, the next location, and the corresponding time in the future. Machine learning methods include supervised, semi-supervised and unsupervised learning, and mainly make the classification or regression of inputs.

Joseph *et al.* [53] propose a Bayesian nonparametric approach to model human mobility patterns and solve issues of over-fitting or under-fitting. Song *et al.* [100] design an intelligent system for predicting human mobility in urban scenarios based on the deep learning architecture from multisource heterogeneous data. Monreale *et al.* [71] propose an effective method to predict the next location of people using decision tree algorithm. Song *et al.* [99] design an intelligent system (DeepTransport) to predict people's future movements and travel conditions based on the deep learning architecture. Feng *et al.* [34] propose an attentional recurrent network for mobility prediction from lengthy and sparse trajectories and outperform the state-of-the-art models by a series of experiments. Baumann *et al.* [13] predict human mobility by choosing the most appropriate model in the field of machine learning and demonstrate that different machine learning methods for different periods can greatly improve the performance. Kun *et al.* [78] develop a novel non-parametric generative model for location trajectories that mines the statistical features of human mobility.

# 6. CHALLENGES

In previous sections, we have surveyed the state of the art related to human mobility models and prediction methods. We analyze the characteristics of individual and collective mobility models and describe different types of human mobility prediction methods. There are still many open issues that need to be resolved. In this section, we further outline some challenges, which shed light on the study of urban human mobility.

## 6.1 Data Acquisition

Data acquisition techniques can continually collect data in urban scenarios and play an important role in modeling human mobility. The traditional methods mainly leverage population census and travel questionnaire to obtain people's travel information and are often costly and inefficient. The popularity of handheld and GPS devices makes it easier to obtain trajectory data for human. Human can be as a new sensor to sense the pulse of urban cities. Their travel routes and blog posts contribute to understanding the events happening around them. However, the crowdsourcing approach also brings many challenges.

Smartphones usually need the supply of energy during the sensing process. So it cannot record relative information in the case of insufficient battery power, giving rise to data-missing problems. Meanwhile, people usually have a skewed distribution in some places, making the acquired data more sparse. In addition, the sample of data is based on partial users and is still insufficient to characterize the population level of human mobility. Therefore, it is necessary to design novel data acquisition technologies to collect multi-source data in the big data era.

## 6.2 Data Analytics

Multi-source heterogeneous data provides a great opportunity for uncovering urban human mobility patterns, and also presents new challenges to analyze the characteristics of human mobility. Some individual and population mobility models have been proposed to capture the laws of people's movement in urban scenarios. Meanwhile, 4 types of prediction methods have been surveyed to show the predictability and exploration of human mobility. However, the above-mentioned models and methods are utilized to deal with one single dataset. The data analytics based on multi-source datasets still needs to be deeply studied. Data fusion is not simply adding together a series of features extracted from multi-source datasets but requires a full understanding of each dataset and efficient utilization them from all aspects of a data analysis framework.

The huge volume of human mobility data makes it possible to predict the future migration flow between origins and destinations and the next locations of people in their daily life. However, trajectory data is not completely accurate owing to device and human errors [132]. So it is vital to delete such data noise from multi-source data before starting a data analysis task. Consequently, we need to use an appropriate map matching algorithm to transform latitude and longitude coordinates into relevant geographical location.

## 6.3 Data Privacy

GPS devices, smartphones, and social media, which are utilized to record people's locations, time, and social activities and hence include the users' private information. Most of the models and algorithms for predicting individual mobility require participants to share their historical travel information. In addition, some blogs on social media possess users' photos, events, and social relationship. Therefore, the privacy issue for people providing personal information becomes extremely important [63].

The trajectory data contains daily travel routes. Attackers may use the spatial and temporal correlations hidden in a user's trajectory data to deduce their mobility patterns and identify their home and workplaces. A bunch of techniques have been proposed to protect users' trajectory data from privacy leak. The privacy technology encrypts users' location information while assuring effective use of trajectory data. Specifically, existing privacy-preserving technologies include clustering-based [1], generalization-based [7], suppression-based [106], and grid-based approach [39].

## 6.4 Prediction Accuracy

Urban human mobility prediction mainly focuses on the estimation of the time-bin and the next location that a person will visit in a city. The prediction accuracy in existing literature lies between 40% and 93%. Furthermore, researchers verify that spatiotemporal resolution of data has an important impact on the prediction accuracy of algorithms. In addition, human mobility possesses the exploration of new places, also influencing the prediction accuracy. Cuttone *ea al.* [24] demonstrate that 20-25% of destinations are places that have never been visited, and 70% of places are visited only once. Therefore, exploration brings challenges for predicting human mobility in the future.

A series of research on human mobility prediction mainly focuses on regular location prediction but ignores people's exploration problem. We still lack an extremely accurate method for dealing with this task in urban cities. Because all the models are based on a set of historical visited places and cannot predict a new and unexpected location. Other regularities hidden in trajectory data may be useful for improving the prediction accuracy. For example, we may combine social relationship, POI, social media, and email with spatiotemporal characteristics to achieve a higher precision.

## 7. CONCLUSION

In this survey, we provided an in-depth literature review of recent studies on human mobility in urban scenarios. Specifically, we introduced a variety of heterogeneous data and compare their advantages and disadvantages. Then we described the characteristics and metrics of human mobility and contribute to modeling human mobility. In addition, we reviewed the state of the art in the field of human mobility models and illustrated their characteristics and limitations. Furthermore, we explored four types of human mobility prediction methods and discussed their principles. Finally, we discussed serval challenges and future trends. In particular, a ubiquitous occurrence of mobile devices and social media makes human behaviors more closely linked to social networks, maybe sparking a novel driving force for modeling and predicting human mobility. We hope that this survey can make researchers to acquire a deeper understanding of the literature of human mobility and arouse new research interests in this field.

## 8. ACKNOWLEDGEMENTS

This work was partially supported by the National Natural Science Foundation of China (61572106), the Natural Science Foundation of Liaoning Province, China (201602154), and the Fundamental Research Funds for the Central Universities (DUT18JC09).

## 9. REFERENCES

[1] O. Abul, F. Bonchi, and M. Nanni. Never walk alone: Uncertainty for anonymity in moving objects databases. In *IEEE International Conference on Data Engineering*, pages 376–385, 2008.

[2] M. R. Alam, M. B. I. Reaz, and M. A. M. Ali. Speed: An inhabitant activity prediction algorithm for smart homes. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 42(4):985–990, 2012.

[3] L. Alessandretti, P. Sapiezynski, S. Lehmann, and A. Baronchelli. Multi-scale spatio-temporal analysis of human mobility. *Plos One*, 12(2):e0171686, 2017.

[4] L. Alessandretti, P. Sapiezynski, V. Sekara, S. Lehmann, and A. Baronchelli. Evidence for a conserved quantity in human mobility. *Nature Human Behaviour*, 2:485–491.

[5] J. E. Anderson. The gravity model. *Nber Working Papers*, 19(3):979–981, 2011.

[6] N. Aschenbruck, A. Munjal, and T. Camp. Trace-based mobility modeling for multi-hop wireless networks. *Computer Communications*, 34(6):704–714, 2011.

[7] M. Atzori, M. Atzori, and Y. Saygin. Towards trajectory anonymization: a generalization-based approach. In *Sigspatial ACM Gis 2008 International Workshop on Security and Privacy in Gis and Lbs*, pages 52–61, 2008.

[8] D. Balcan, V. Colizza, B. Gonálves, H. Hu, J. Ramasco, and A. Vespignani. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences of the United States of America*, 106(51):21484–21489, 2009.

[9] D. Balcan and A. Vespignani. Phase transitions in contagion processes mediated by recurrent mobility patterns. *Nature Physics*, 7(7):581–586, 2011.

[10] H. Barbosa, M. Barthelemy, G. Ghoshal, C. R. James, M. Lenormand, T. Louail, R. Menezes, J. J. Ramasco, F. Simini, and M. Tomasini. Human mobility: Models and applications. *Physics Reports*, 734:1 – 74, 2018.

[11] G. Barlacchi, M. D. Nadai, R. Larcher, A. Casella, C. Chitic, G. Torrisi, F. Antonelli, A. Vespignani, A. Pentland, and B. Lepri. A multi-source dataset of urban life in the city of milan and the province of trentino. *Scientific Data*, 2:150055, 2015.

[12] M. Barthélemy. Spatial networks. *Physics Reports*, 499(1):1–101, 2011.

[13] P. Baumann, C. Koehler, A. Dey, and S. Santini. Selecting individual and population models for predicting human mobility. *IEEE Transactions on Mobile Computing*, 17(10):2408–2422, 2018.

[14] A. Bhattacharya and S. K. Das. Lezi-update: an information-theoretic approach to track mobile users in pcs networks. pages 1–12, 1999.

[15] J. I. Blanford, Z. Huang, A. Savelyev, and A. M. Maceachren. Geo-located tweets. enhancing mobility maps and capturing cross-border movement. *Plos One*, 10(6):e0129202, 2015.

[16] P. Bonacich. Some unique properties of eigenvector centrality. *Social Networks*, 29(4):555–564, 2007.

[17] G. Box and G. M. Jenkins. *Box and Jenkins: Time Series Analysis, Forecasting and Control*, pages 161–215. Palgrave Macmillan UK, Basingstoke, Hampshire, 2013.

[18] D. Brockmann, L. Hufnagel, and T. Geisel. The scaling laws of human travel. *Nature*, 439(7075):462–465, 2006.

[19] F. Calabrese, M. Diao, G. Di Lorenzo, J. Ferreira, and C. Ratti. Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transportation research part C: emerging technologies*, 26:301–313, 2013.

[20] F. Calabrese, L. Ferrari, and V. D. Blondel. Urban sensing using mobile phone network data: a survey of research. *Acm computing surveys (csur)*, 47(2):25, 2015.

[21] B. B. S. Chen. Modeling destination choice in hurricane evacuation with an intervening opportunity model. *Surface & Coatings Technology*, 213(12):48–58, 2005.

[22] D. J. Crandall, L. Backstrom, D. Cosley, S. Suri, D. Huttenlocher, and J. Kleinberg. Inferring social ties from geographic coincidences. *Proceedings of the National Academy of Sciences*, 107(52):22436–22441, 2010.

[23] B. C. Csáji, A. Browet, V. A. Traag, J. C. Delvenne, E. Huens, P. Van Dooren, Z. Smoreda, and V. D. Blondel. Exploring the mobility of mobile phone users. *Physica A: Statistical Mechanics and its Applications*, 392(6):1459–1473, 2013.

[24] A. Cuttone, S. Lehmann, and M. C. González. Understanding predictability and exploration in human mobility. *Epj Data Science*, 7(1):2, 2018.

[25] E. M. Daly and M. Haahr. Social network analysis for information flow in disconnected delay-tolerant manets. *IEEE Transactions on Mobile Computing*, 8(5):606–621, 2009.

[26] F. N. David. Hypothesis-testing in time series analysis. *Biometrika*, 39(265):300–325, 1952.

[27] M. De Domenico, A. Lima, and M. Musolesi. Interdependence and predictability of human mobility and social interactions. *Pervasive & Mobile Computing*, 9(6):798–807, 2012.

[28] P. Deville, C. Song, N. Eagle, V. D. Blondel, A. L. Barabási, and D. Wang. Scaling identity connects human mobility and social interactions. *Proceedings of the National Academy of Sciences*, 113(26):7047–7052, 2016.

[29] M. D. Domenico, A. Lima, M. C. González, and A. Arenas. Personalized routing for multitudes in smart cities. *Epj Data Science*, 4(1):1, 2015.

[30] J. L. Doob. The brownian movement and stochastic equations. *Annals of Mathematics*, 43(2):351–369, 1942.

[31] B. Du, C. Liu, W. Zhou, Z. Hou, and H. Xiong. Catch me if you can: Detecting pickpocket suspects from large-scale transit records. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 87–96, San Francisco, 2016.

[32] A. Einstein. Investigations on the theory of the brownian movement. 7(10):281, 1956.

[33] S. Eubank, H. Guclu, V. S. Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang. Modelling disease outbreaks in realistic urban social networks. *Nature*, 429(6988):180–184, 2004.

[34] J. Feng, Y. Li, C. Zhang, F. Sun, F. Meng, A. Guo, and D. Jin. Deepmove: Predicting human mobility with attentional recurrent networks. In *In Proceedings of the 2018 World Wide Web Conference on World Wide Web*, pages 1459–1468, Lyon, 2018.

[35] R. Feynman. The brownian movement. *The Feynman Lectures of Physics*, 1:41–1, 1964.

[36] R. Flowerdew and M. Aitkin. A method of fitting the gravity model based on the poisson distribution. *J Reg Sci*, 22(2):191–202, 1982.

[37] L. C. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 40(1):35–41, 1977.

[38] L. C. Freeman. Centrality in social networks conceptual clarification. *Social Networks*, 1(3):215–239, 2008.

[39] G. Gidofalvi, X. Huang, and T. B. Pedersen. Privacy-preserving data mining on moving object trajectories. In *International Conference on Mobile Data Management*, pages 60–68, Mannheim, 2007.

[40] M. C. Gonzalez, C. A. Hidalgo, and A. L. Barabasi. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, 2008.

[41] K. Gopalratnam and D. J. Cook. Online sequential prediction via incremental parsing: The active lezi algorithm. *IEEE Intelligent Systems*, 22(1):52–58, 2007.

[42] X. P. Han, Q. Hao, B. H. Wang, and T. Zhou. Origin of the scaling law in human mobility: hierarchy of traffic systems. *Physical Review E Statistical Nonlinear & Soft Matter Physics*, 83(2):036117, 2011.

[43] A. Hess, K. A. Hummel, and W. N. Gansterer. Data-driven human mobility modeling:a survey and engineering guidance for mobile networking. *Acm Computing Surveys*, 48(3):1–39, 2015.

[44] L. Hong, Y. Zheng, D. Yung, J. Shang, and L. Zou. Detecting urban black holes based on human mobility data. In *Sigspatial International Conference on Advances in Geographic Information Systems*, page 35, 2015.

[45] R. Horak. *Telecommunications and Data Communications Handbook*. Wiley-Interscience, New Jersey, 2012.

[46] H. Hsieh, S. Lin, and Y. Zheng. Inferring air quality for station location recommendation based on urban big data. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 437–446, Sydney, 2015.

[47] H. Hsieh, R. Yan, and C. T. Li. Where you go reveals who you know: Analyzing social ties from millions of footprints. In *ACM International on Conference on Information and Knowledge Management*, pages 1839–1842, 2015.

[48] M. Itoh, D. Yokoyama, M. Toyoda, Y. Tomita, S. Kawamura, and M. Kitsuregawa. Visual exploration of changes in passenger flows and tweets on mega-city metro network. *IEEE Transactions on Big Data*, 2(1):85–99, 2016.

[49] B. Jedari, F. Xia, and Z. Ning. A survey on human-centric communications in non-cooperative wireless relay networks. *IEEE Communications Surveys & Tutorials*, pages 1–9, 2018.

[50] J. S. Jia, J. Jia, C. K. Hsee, and B. Shiv. The role of hedonic behavior in reducing perceived risk: Evidence from postearthquake mobile-app data. *Psychological Science*, 28(1):23–35, 2016.

[51] B. Jiang, J. Yin, and S. Zhao. Characterizing the human mobility pattern in a large street network. *Physical Review E Statistical Nonlinear and Soft Matter Physics*, 80(2 Pt 1):021136, 2009.

[52] S. Jiang, J. Ferreira, and M. C. Gonzalez. Activity-based human mobility patterns inferred from mobile phone data: A case study of singapore. *IEEE Transactions on Big Data*, 3(2):208–219, 2017.

[53] J. Joseph, F. Doshi-Velez, and N. Roy. A bayesian non-parametric approach to modeling mobility patterns. In *Twenty-Fourth AAAI Conference on Artificial Intelligence*, pages 1587–1593, Atlanta, 2010.

[54] W. S. Jung. Gravity model in the korean highway. 81(4):48005, 2008.

[55] J. M. Kleinberg. Navigation in a small world. *Nature*, 406(6798):845, 2000.

[56] 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*, 2018. DOI:10.1007/s11280-018-0578-x.

[57] X. 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*, 21(3):825–847, 2018.

[58] X. Kong, F. Xia, Z. Ning, A. Rahim, Y. Cai, Z. Gao, and J. Ma. Mobility dataset generation for vehicular social networks based on floating car data. *IEEE Transactions on Vehicular Technology*, 67(5):3874–3886, 2018.

[59] X. Kong, F. Xia, J. Wang, A. Rahim, and S. K. Das. Time-location-relationship combined service recommendation based on taxi trajectory data. *IEEE Transactions on Industrial Informatics*, 13(3):1202–1212, 2017.

[60] X. Kong, Z. Xu, G. Shen, J. Wang, Q. Yang, and B. Zhang. Urban traffic congestion estimation and prediction based on floating car trajectory data. *Future Generation Computer Systems*, 61:97–107, 2016.

[61] G. Krings, F. Calabrese, C. Ratti, and V. D. Blondel. Urban gravity: a model for intercity telecommunication flows. *Journal of Statistical Mechanics Theory and Experiment*, 2009(7):L07003, 2009.

[62] M. K. Le, A. Bhaskar, and E. Chung. Passenger segmentation using smart card data. *IEEE Transactions on Intelligent Transportation Systems*, 16(3):1537–1548, 2015.

[63] W. C. Lee, J. Krumm, K. Deng, K. Xie, K. Zheng, X. Zhou, G. Trajcevski, C. Y. Chow, M. F. Mokbel, and H. Jeung. Computing with spatial trajectories. *Computing with Spatial Trajectories*, 2011.

[64] M. Lenormand, A. Bassolas, and J. J. Ramasco. Systematic comparison of trip distribution laws and models. *Journal of Transport Geography*, 51:158–169, 2016.

[65] Q. Li, Y. Zheng, X. Xie, Y. Chen, W. Liu, and W. Y. Ma. Mining user similarity based on location history. In *ACM Sigspatial International Conference on Advances in Geographic Information Systems*, page 34. ACM, 2008.

[66] X. Liang, J. Zhao, D. Li, and K. Xu. Unraveling the origin of exponential law in intra-urban human mobility. *Scientific Reports*, 3(10):65–65, 2013.

[67] X. Lu, L. Bengtsson, and P. Holme. Predictability of population displacement after the 2010 haiti earthquake. *Proceedings of the National Academy of Sciences of the United States of America*, 109(29):11576–11581, 2012.

[68] M. K. E. Mahrsi, E. Côme, L. Oukhellou, and M. Verleysen. Clustering smart card data for urban mobility analysis. *IEEE Transactions on Intelligent Transportation Systems*, 18(3):712–728, 2017.

[69] A. P. Masucci, J. Serras, A. Johansson, and M. Batty. Gravity versus radiation models: on the importance of scale and heterogeneity in commuting flows. *Physical Review E Statistical Nonlinear & Soft Matter Physics*, 88(2):022812, 2013.

[70] L. Matyas. Proper econometric specification of the gravity model. *World Economy*, 20(3):363–368, 1997.

[71] A. Monreale, F. Pinelli, R. Trasarti, and F. Giannotti. Wherenext: A location predictor on trajectory pattern mining. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 637–646, New York, 2009. ACM.

[72] E. W. Montroll. Random walks on lattices. In *II J. Math. Phys*, pages 193–220, 1969.

[73] M. E. J. Newman. *Networks: an introduction*. Oxford University Press, Inc., Oxford, 2010.

[74] M. Ni, Q. He, and J. Gao. Forecasting the subway passenger flow under event occurrences with social media. *IEEE Transactions on Intelligent Transportation Systems*, 18(6):1623–1632, 2017.

[75] Z. Ning, F. Xia, N. Ullah, X. Kong, and X. Hu. Vehicular social networks: Enabling smart mobility. *IEEE Communications Magazine*, 55(5):16–55, 2017.

[76] A. Noulas, S. Scellato, R. Lambiotte, M. Pontil, and C. Mascolo. A tale of many cities: universal patterns in human urban mobility. *Plos One*, 7(9):e37027, 2012.

[77] J. D. D. Ortúzar and L. G. Willumsen. Modeling transport. *Transactions of the Institute of British Geographers*, 18(1), 2011.

[78] K. Ouyang, R. Shokri, D. S. Rosenblum, and W. Yang. A non-parametric generative model for human trajectories. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, pages 3812–3817, Stockholm, 2018.

[79] V. Palchykov, M. Mitrović, H. H. Jo, J. Saramäki, and R. K. Pan. Inferring human mobility using communication patterns. *Sci Rep*, 4:6174, 2014.

[80] J. R. B. Palmer, T. J. Espenshade, F. Bartumeus, Y. C. Chang, N. E. Ozgencil, and K. Li. New approaches to human mobility: Using mobile phones for demographic research. *Demography*, 50(3):1105–1128, 2013.

[81] K. Pearson. The problem of the random walk. *Nature*, 72:342, 1905.

[82] K. Peterson. Business telecom systems: A guide to choosing the best technologies and services. *Crc Press*, page 412, 2000.

[83] H. Pham, C. Shahabi, and Y. Liu. Ebm: an entropy-based model to infer social strength from spatiotemporal data. In *ACM SIGMOD International Conference on Management of Data*, pages 265–276, 2013.

[84] C. Potter. New questions in the 1940 census. *Prologue-Quarterly of the National Archives and Records Administration*, 42(4):46–52, 2010.

[85] S. K. Pulliyakode and S. Kalyani. A modified ppm algorithm for online sequence prediction using short data records. *Communications Letters IEEE*, 19(3):423–426, 2015.

[86] S. Qiao, D. Shen, X. Wang, N. Han, and W. Zhu. A self-adaptive parameter selection trajectory prediction approach via hidden markov models. *IEEE Transactions on Intelligent Transportation Systems*, 16(1):284–296, 2015.

[87] Y. Qiao, Z. Si, Y. Zhang, F. B. Abdesslem, X. Zhang, J. Yang, Y. Qiao, Z. Si, Y. Zhang, and F. B. Abdesslem. A hybrid markov-based model for human mobility prediction. *Neurocomputing*, 278:99–109, 2018.

[88] A. Rahim, T. Qiu, Z. Ning, J. Wang, N. Ullah, A. Tolba, and F. Xia. Social acquaintance based routing in vehicular social networks. *Future Generation Computer Systems*, 2017.

[89] Y. Ren, M. Ercseyravasz, P. Wang, M. C. González, and Z. Toroczkai. Predicting commuter flows in spatial networks using a radiation model based on temporal ranges. *Nature Communications*, 5(5):5347, 2014.

[90] I. Rhee, M. Shin, S. Hong, K. Lee, S. J. Kim, and S. Chong. On the levy-walk nature of human mobility. *IEEE/ACM Transactions on Networking*, 19(3):630–643, 2011.

[91] E. R. Ruiter. Toward a better understanding of the intervening opportunities model. *Transportation Research*, 1(1):47–56, 1967.

[92] A. Rutherford, M. Cebrian, S. Dsouza, E. Moro, A. Pentland, and I. Rahwan. Limits of social mobilization. *Proceedings of the National Academy of Sciences of the United States of America*, 110(16):6281–6, 2013.

[93] S. Scellato, A. Noulas, and R. Lambiotte. Socio-spatial properties of online location-based social networks. *Proceedings of Icwsm*, page 2011, 2012.

[94] F. Simini, M. C. González, A. Maritan, and A. L. Barabási. A universal model for mobility and migration patterns. *Nature*, 484(7392):96–100, 2012.

[95] F. Simini and A. Maritan. Human mobility in a continuum approach. *Plos One*, 8(3):e60069, 2013.

[96] C. Song, T. Koren, P. Wang, and A. L. Barabási. Modelling the scaling properties of human mobility. *Nature Physics*, 6(10):818–823, 2010.

[97] C. Song, Z. Qu, N. Blumm, and A. L. Barabási. Limits of predictability in human mobility. *Science*, 327(5968):1018–1021, 2010.

[98] L. Song, D. Kotz, R. Jain, and X. He. Evaluating next-cell predictors with extensive wi-fi mobility data. *IEEE Transactions on Mobile Computing*, 5(12):1633–1649, 2006.

[99] X. Song, H. Kanasugi, and R. Shibasaki. Deeptransport: prediction and simulation of human mobility and transportation mode at a citywide level. In *International Joint Conference on Artificial Intelligence*, pages 2618–2624, New York, 2016.

[100] X. Song, H. Kanasugi, and R. Shibasaki. Deeptransport: Prediction and simulation of human mobility and transportation mode at a citywide level. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence*, pages 2618–2624, New York, 2016. International Joint Conferences on Artificial Intelligence Organization.

[101] X. Song, Q. Zhang, Y. Sekimoto, R. Shibasaki, N. J. Yuan, and X. Xie. Prediction and simulation of human mobility following natural disasters. *ACM Trans. Intell. Syst. Technol.*, 8(2):29:1–29:23, 2016.

[102] S. A. Stouffer. Intervening opportunities: A theory relating mobility and distance. *American Sociological Review*, 5(6):845–867, 1940.

[103] L. Sun and K. W. Axhausen. Understanding urban mobility patterns with a probabilistic tensor factorization framework. *Transportation Research Part B: Methodological*, 91:511–524, 2016.

[104] L. Sun, K. W. Axhausen, D.-H. Lee, and X. Huang. Understanding metropolitan patterns of daily encounters. *Proceedings of the National Academy of Sciences*, 110(34):13774–13779, 2013.

[105] F. Terroso-Sáenz, J. Cuenca-Jara, A. González-Vidal, and A. F. Skarmeta. Human mobility prediction based on social media with complex event processing. *International Journal of Distributed Sensor Networks*, 12(9):5836392–5836392, 2016.

[106] M. Terrovitis and N. Mamoulis. Privacy preservation in the publication of trajectories. In *International Conference on Mobile Data Management*, pages 65–72, Peking, 2008.

[107] M. Tizzoni, P. Bajardi, A. Decuyper, K. K. G. Kon, C. M. Schneider, V. Blondel, Z. Smoreda, M. C. González, and V. Colizza. On the use of human mobility proxies for modeling epidemics. *Plos Computational Biology*, 10(7):e1003716, 2014.

[108] E. Toch, B. Lerner, E. Ben-Zion, and I. Ben-Gal. Analyzing large-scale human mobility data: a survey of machine learning methods and applications. *Knowledge and Information Systems*, 2018. DOI:10.1007/s10115-018-1186-x.

[109] H. Tong, T. K. Kumar, and Y. Huang. *Stationary Time Series Models*. Wiley-Blackwell, New Jersey, 2011.

[110] J. Treurniet. A taxonomy and survey of microscopic mobility models from the mobile networking domain. *Acm Computing Surveys*, 47(1):1–32, 2014.

[111] M. Veloso, S. Phithakkitnukoon, C. Bento, N. Fonseca, and P. Olivier. Exploratory study of urban flow using taxi traces. In *First Workshop on Pervasive Urban Applications (PURBA)*, pages 241–248. Pervasive Computing, 2011.

[112] G. Walker. On periodicity in series of related terms. *Proceedings of the Royal Society of London*, 131(818):518–532, 1931.

[113] J. Wang, X. Kong, A. Rahim, F. Xia, A. Tolba, and Z. Al-Makhadmeh. Is2fun: Identification of subway station functions using massive urban data. *IEEE Access*, 5:27103–27113, 2017.

[114] W. Wang, L. Pan, N. Yuan, S. Zhang, and D. Liu. A comparative analysis of intra-city human mobility by taxi. *Physica A: Statistical Mechanics and its Applications*, 420:134–147, 2015.

[115] Z. Wang, Y. Zhang, Y. Li, Q. Wang, and F. Xia. Exploiting social influence for context-aware event recommendation in event-based social networks. In *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, pages 1–9, 2017.

[116] A. G. Wilson. *Urban and regional models in geography and planning*. Wiley,, New Jersey, 1974.

[117] F. Xia, L. Liu, B. Jedari, and S. K. Das. Pis: A multi-dimensional routing protocol for socially-aware networking. *IEEE Transactions on Mobile Computing*, 15(11):2825–2836, 2016.

[118] F. Xia, L. Liu, J. Li, A. M. Ahmed, L. T. Yang, and J. Ma. Beeinfo: Interest-based forwarding using artificial bee colony for socially aware networking. *IEEE Transactions on Vehicular Technology*, 64(3):1188–1200, 2015.

[119] F. Xia, L. Liu, J. Li, J. Ma, and A. V. Vasilakos. Socially aware networking: A survey. *IEEE Systems Journal*, 9(3):904–921, 2017.

[120] F. Xia, A. Rahim, X. Kong, M. Wang, Y. Cai, and J. Wang. Modeling and analysis of large-scale urban mobility for green transportation. *IEEE Transactions on Industrial Informatics*, 14(4):1469–1481, 2018.

[121] F. Xia, J. Wang, X. Kong, Z. Wang, J. Li, and C. Liu. Exploring human mobility patterns in urban scenarios: A trajectory data perspective. *IEEE Communications Magazine*, 56(3):142–149, 2018.

[122] X. Y. Yan, W. X. Wang, Z. Y. Gao, and Y. C. Lai. Universal model of individual and population mobility on diverse spatial scales. *Nature Communications*, 8(1), 2017.

[123] Y. Yang, N. V. Chawla, P. Basu, B. Prabhala, and T. L. Porta. Link prediction in human mobility networks. In *Ieee/acm International Conference on Advances in Social Networks Analysis and Mining*, pages 380–387, 2013.

[124] F. Yi, Y. Gu, H. Wang, Y. Sun, and L. Sun. Mobility pattern based relationship inference from spatiotemporal data. In *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, pages 1–6, 2017.

[125] M. Yin, M. Sheehan, S. Feygin, J. Paiement, and A. Pozdnoukhov. A generative model of urban activities from cellular data. *IEEE Transactions on Intelligent Transportation Systems*, 19(6):1682–1696, 2018.

[126] Z. Yu, H. Wang, B. Guo, T. Gu, and T. Mei. Supporting serendipitous social interaction using human mobility prediction. *IEEE Transactions on Human-Machine Systems*, 45(6):811–818, 2015.

[127] G. U. Yule. On a method of investigating periodicities in disturbed series, with special reference to wolfer's sunspot numbers. *Philosophical Transactions of the Royal Society of London*, 226(226):267–298, 1927.

[128] S. Zeng, H. Wang, Y. Li, and D. Jin. Predictability and prediction of human mobility based on application-collected location data. In *IEEE International Conference on Mobile Ad Hoc and Sensor Systems*, pages 28–36, 2017.

[129] F. Zhang, J. Zhao, C. Tian, C. Xu, X. Liu, and L. Rao. Spatiotemporal segmentation of metro trips using smart card data. *IEEE Transactions on Vehicular Technology*, 65(3):1137–1149, 2016.

[130] J. Zhao, F. Zhang, L. Tu, C. Xu, D. Shen, C. Tian, X. Y. Li, and Z. Li. Estimation of passenger route choice pattern using smart card data for complex metro systems. *IEEE Transactions on Intelligent Transportation Systems*, 18(4):790–801, 2017.

[131] K. Zhao, M. Musolesi, P. Hui, W. Rao, and S. Tarkoma. Explaining the power-law distribution of human mobility through transportation modality decomposition. *Scientific reports*, 5(9136):1–7, 2015.

[132] Y. Zheng. *Trajectory Data Mining: An Overview*. ACM, New York, 2015.

[133] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W. Y. Ma. Understanding mobility based on gps data. In *International Conference on Ubiquitous Computing*, pages 312–321. ACM, 2008.

[134] Y. Zheng, X. Xie, and W. Y. Ma. Geolife: A collaborative social networking service among user, location and trajectory. *Bulletin of the Technical Committee on Data Engineering*, 33(2):32–39, 2011.

[135] Y. Zheng, L. Zhang, X. Xie, and W. Y. Ma. Mining interesting locations and travel sequences from gps trajectories. In *International Conference on World Wide Web*, pages 791–800. ACM, 2009.

[136] G. K. Zipf. The p1p2/d hypothesis: On the intercity movement of persons. *American Sociological Review*, 11(6):677–686, 1946.

[137] J. Ziv and A. Lempel. *Compression of individual sequences via variable-rate coding*. IEEE Press, New Jersey, 1978.