A

Human Mobility Modeling Methodologies and Applications: An Overview

YI DING, University of Minnesota

As a basic issue for many applications related to smart cities and urban computing, studies on human mobility undergo rapid proliferation in the recent years. Thanks to the pervasive mobile phones and sensor networks and the fast developed Internet, massive data sets are collected so that we can extract the human mobility pattern from the data. Based on the human mobility pattern/models, plenty of applications can be achieved such as transportation forecasting, urban planning and crowdsourcing. In this survey, the recently emerged studies on human mobility are reviewed and categorized. Some early but classical studies are also referred for completeness. We will focus on the modeling methodologies and applications of human mobility studies. Some other related topics such as data sets will also be discussed.

Additional Key Words and Phrases: Human mobility, smart cities, urban computing

1. INTRODUCTION

Cities around the world are currently under quick transition towards a low carbon environment, high quality of living, and resource efficient economy. Smart cities and urban computing are the two related research fields both committed to this goal. Among all the topics in the field, the human mobility is the major concern and basis in many studies such as infrastructure construction, transportation, and epidemics. Understanding human mobility can substantially enhance our comprehension and control on urban traffic [Pan et al. 2013; Chen et al. 2015], urban planning [Zheng et al. 2011; Yuan et al. 2012], mobile network routing [Shin et al. 2008; Zhao et al. 2015a] and epidemics [Colizza et al. 2007; Belik et al. 2011].

In this article, we survey recent literature on urban mobility and focused more on modeling methodologies and applications. Early work on human mobility usually tries to build a generic model [Brockmann et al. 2006; Gonzalez et al. 2008]. However, with the fast development of Internet and mobile phones, more and more studies tend to be application-oriented. Even the modeling and prediction methods are customized for the specific application scenario such as balancing in bike-sharing system [Yang et al. 2016] and large group dynamic sensing in religious gathering [Jamil et al. 2015]. That is also the reason that this survey place emphasis on studies of application of human mobility.

As an evolving research field, few survey papers have been put forward on human mobility. [Zhao et al. 2016] provides a relatively complete review of many studies on human mobility to date. The recent studies are summarized from a data mining view following a knowledge discovery process. Another high quality survey reviews the existing works on human mobility but the applications are focused exclusively on mobile networking [Hess et al. 2016]. The related surveys on human mobility will be talked in details in Section II

Although the literature on human mobility varies in many aspects, some basic concepts such as trip should be clarified in the first place. Before we talk about the modeling and the prediction of human mobility, the predictability of human mobility should also be discussed. These basic background knowledges will be presented in Section III

Data plays an important role in the studies on human mobility. Except for some early synthetic work [Bettstetter et al. 2003], most recent work relies on certain data mining technology to extract the human mobility from the raw data. Widely used data sets include mobile phone records [Gonzalez et al. 2008; Song et al. 2010; Isaacman et al. 2012], taxicab data [Jiang et al. 2009; Ganti et al. 2013] and Automatic Fare Collection (AFC) data [Lathia and Capra 2011; Bhattacharya et al. 2013]. Some large cities such as New York City are holding data disclosure project expecting innovative research and applications. A detailed summary of data sets used in human mobility studies as well as some privacy issues can be found in Section IV.

The human mobility models are the bridges from data to the application. Gravity law tries to find the relation between crowd movement and the city population and distance, a similar idea of Newton's law of universal gravitation [Erlander and Stewart 1990]. Levy flight is a widely used model to mimic human mobility as a kind of random walk [Brockmann et al. 2006; Shin et al. 2008; Jiang et al. 2009]. Human mobility is shown to have more regularity than we thought. Radiation model is another effective model when previous data is lack [Simini et al. 2012]. Compared with tradition model that is built based on single source data set, a multi-view model is shown to be more accurate [Zhang et al. 2014]. Activity-based model analyzes the mechanism underlying the group movement [Jiang et al. 2016]. The recent human mobility models are summarized in Section V.

Applications of the human mobility comes in several main categorizes. Traffic forecasting is the most direct and widely application. Specifically, the traffic forecasting can be divided in to crowd/group forecasting [Li et al. 2012; Pan et al. 2013; Jamil et al. 2015; Chen et al. 2015; Fan et al. 2015; Shimosaka et al. 2015; Zhang et al. 2017] and individual mobility prediction [Zhang et al. 2015a; Song et al. 2016; Kulkarni et al. 2016]. Other applications include urban planning [Zheng et al. 2011; Yuan et al. 2012; Jiang et al. 2016], carpooling [Trasarti et al. 2011], crowdsourcing [He et al. 2015], routing in mobile network [Shin et al. 2007; Shin et al. 2008; Zhao et al. 2015a], forecast and control of epidemics [Hufnagel et al. 2004; Colizza et al. 2007; Belik et al. 2011], and other mobile and network application [Zheng et al. 2013; Yang et al. 2016]. The detailed discussion of a variety of applications based on human mobility can be found in Section VI.

Despite the abundant existing studies on human mobility, some issues remain unsolved for the following researchers. These open issues will be briefly discussed in Section VII.

2. RELATED SURVEY PAPERS

Although the study on human mobility can even date back to the 1940s [Zipf 1946], the majority of research in this field came out in the very recent years thanks to the developed Internet and ubiquitous sensor network. There are some survey papers trying to characterize the existing research on human mobility from different angles. In [Asgari et al. 2013], the author reviewed the major characteristics of human mobility. A well-organized summary of the factors that influence individual mobility patters is provided in the survey. However, the classification of human mobility studies into three categories: trajectory-based studies, dynamic proximity networks, and flow on networks is not well defined and clarified. Besides, the data collection techniques summarized in the survey only covers cellular network and Wi-Fi network, many other techniques like transportation data and social network data are not covered in the survey. In [Hess et al. 2016], a variety of mobility models are reviewed and classified according to the features of the model. A well depicted figure and a clearly organized table are presented to summarize the existing mobility models. However, this survey is focused exclusively on the application of mobile network, studies and applications on transportation and social network are not reviewed. In [Zhao et al. 2016], recent studies and progress in urban human mobility data mining are reviewed. The related studies are classified according to their function in the knowledge discovery process based on data mining and machine learning. The studies reviewed in the survey are up-to-date but not very complete due to the page limits.

Compared to the surveys aforementioned, this survey focuses more on the models and applications, as more and more studies on human mobility tend to be application-oriented. We also introduce some basic concepts and remaining questions in this field, which can benefit the following researchers.

3. BASIC CONCEPTS

Before we begin the discussion on a variety of modeling methods of human mobility, a question we may raise is to what extent the human mobility is predictable. Strikingly, research in [Song et al. 2010] shows that the human mobility is 93% predictable despite the apparent randomness of the individuals' trajectories. The entropy is calculated and Fano's inequality is used to get a lower bound of the error probability.

Significant temporal and spatial regularities are found in human mobility [Gonzalez et al. 2008; Jiang et al. 2009; Ganti et al. 2013]. That is, human tends to commute between home and work during weekdays and shop and go for entertainments on weekends. Statistically, short distance and short time trips make up most of the human movements. While rare but nonnegligible long distance trips show up in the long tail and heavy tail of the distribution. The underlying reasons are still behind the curtain. Some researchers argue the reason lies in the structure of the road network [Jiang et al. 2009]. While some try to explain the phenomenon by decomposing the transportation modes [Zhao et al. 2015b]. Another point of view is the activity-based analysis [Jiang et al. 2016].

For most data sets, the human mobility is hided in a bunch of continues spatial-temporal trails, with timestamp and location stamp. In order to build further analysis, some common basic concepts are needed. A **trip** or **trajectory** is usually used as a unit in mobility data analyzing, with a start time/position and end time/position. In [Ganti et al. 2013], a graph theory concept is used to identify the start and end location of a trip. In [Trasarti et al. 2011], basic concepts such as trip and stop are also defined. The concept **flight** is also used in many literatures to describe the basic unit for human mobility. The scaling and distribution of the length and duration of the trips/trajectories/flights are the focuses of many research in this area [Zhao et al. 2016].These three words are widely used interchangeably, thus we do not differentiate the differences of these words and use them interchangeably in this survey.

4. DATA SOURCES

Summaries of many typical data sets used in human mobility studies can be found in [Hess et al. 2016; Zhao et al. 2016] with clear classification. Here we also provide a review and a table of the commonly used data sets. The difference between our review and those aforementioned is that the data sets will be well categorized according to the source type. Besides, the data sets reviewed here are more complete and up to date. A table that summarizing the data sets can be found in Table I.

Before the ubiquitous spread of the Internet and sensor network, studies of human mobility can only utilize indirect data source as proxy, such as bank notes [Brockmann et al. 2006] and census data [Jung et al. 2008]. The air flight data is widely used in the studies of epidemics [Hufnagel et al. 2004; Colizza et al. 2007]. These data sets are usually coarse-grained and in a wide geographic range, country level or global level.

In some early studies, human mobility data is collected intentionally for the research usage. Geolife dataset was collected in (Microsoft Research Asia) Geolife project by 182 users in a period of over three years [Zheng et al. 2009; Zhao et al. 2015b]. This dataset recoded a broad range of users outdoor movements, including not only life routines but also some entertainments and sports activities. The Nokia MDC dataset is a public dataset that aims to study smartphone user behavior [Zhao et al. 2015b]. The dataset contains extensive smartphone data of two hundred volunteers in the Lake Geneva region over one and a half years. In order to observe the actual behavior of human movement, some groups even hire volunteers to collect walking data with GPS [Shin et al. 2008]. To collect the data intentionally usually calls for extra effort. Data sets used in recent studies usually are collected for other purpose such as billing.

With the widely usage of mobile phones in the recent decade, massive mobile phone call data can be collected. Raw mobile phone call data play important roles in unveiling the regularity pattern in some classical work [Gonzalez et al. 2008; Song et al. 2010]. Later, the mobile phone data came in with a unified data record form: Call Detail Data (CDR). These data are usually collected by the operators/carriers for the billing purpose, but can be used effectively in the studies to find the human mobility at city level [Isaacman et al. 2012; Jiang et al. 2016]. The limitation of the CDR data is that the localization accuracy depends on the density of cellular towers in the area. Besides, without knowing the real location of the human beings, applications based on CDR data are usually coarse-grained. While in recent studies, fine-grained mobile phone localization can be accessed by recording the GPS log information [Fan et al. 2015; Shimosaka et al. 2015]. Unlike any other data sources, mobile phone data provides the location information where he/she goes, with which more applications can be built.

Data mining and analysis on social networking is an independent research field. The users location information is collected to make better service and recommendations. Studies combining social networking and human mobility are not as much as with other types of data, possibly due to the privacy issue [Cho et al. 2011; Hawelka et al. 2014]. The uniqueness of social networking data is it provides a deep understanding of the relationship between people, which can be a factor that influences the behavior of human mobility.

Among all the data sources related to transportation system, taxi data is most widely used in the studies of human mobility. The reason might be that the data sets can be relatively easy to access and GPS information of taxicabs is a direct proxy of human mobility. Zheng Yu built a bunch of applications based on Beijing taxicab data sets [Yuan et al. 2012; Pan et al. 2013; Zheng et al. 2013; Zhang et al. 2017]. Nowadays, many large cities such as New York City are releasing transportation data for solutions of the transportation issues like traffic congestion and pollution [Dimitriou et al. 2016]. The data sets released can play important roles in human mobility research and even can serve as benchmarks in this area. More data sets on taxi cabs can be found in Table I.

With the development in transportation infrastructures, more and more data can be collected in public transportation such as bike [Yang et al. 2016; Zhang et al. 2017], subway [Lathia and Capra 2011; Zhang et al. 2015a] and bus [Bhattacharya et al. 2013]. Unlike taxi, subway and bus usually carry the majority of commuters in the city thus make up a big share in the whole transportation system. The understanding of subway and bus data may have profound influence in the future steps towards smart cities. As a new member in public transportation, bike-sharing systems are deployed in many large metropolitan areas, thus the bike data would more and more meaningful in future. These location data sets are usually specified with the Automated Fare Collection (AFC) system or the locations of bus/bike stations, which can serve as an in-direct proxy of human mobility. The entrance/exit location/time are usually recorded but other information need to be estimated. The General Transit Feed Specification (GTFS) proposed by Google also serve as an important media to collect different kinds of data in a uniformed format for application developing convenience.

Private cars are important parts in city traffic. Although some studies analyzed the mobility patterns of private cars based on the data sets provided by operators [Giannotti et al. 2011; Trasarti et al. 2011], the data of private cars are quite difficult to access. With recent trends of shared-economy, more and more on ground private cars become Uber and Lyft cars. Currently no studies are published based on Uber or Lyft data, but this might be possible in the near future, which can enrich our understanding of private cars and the whole transportation system.

Most studies published try to build the human mobility model based on single data source, which would bring inevitable errors in the modeling. This can be resolved by a multi-view modeling strategy [Zhang et al. 2014; Zhang et al. 2015b]. Data from different sources are considered at the same time to eliminate the error from each other. With the massive data sets collected from different channels, cross-domain data fusion will be increasingly important in the future [Zheng 2015].

Table I. Summary of data sets used in human mobility studies

TYPES	DATA SETS	SCALE LEVEL	DURATION	LOCALIZATION TECHNOLOGY	RELATED STUDIES
Bank notes	U.S. banks notes	Country	-	Bank location	[Brockmann et al. 2006]
Census	Korean census data	Country	-	-	[Jung et al. 2008]
Airplane	IATA airport data	Global	-	Airport location	[Colizza et al. 2007]
	OAG flight data	Global	-	Airport location	[Hufnagel et al. 2004]
Geolife	Microsoft Geolife data	-	3 years	GPS	[Zheng et al. 2009] [Zhao et al. 2015b]
Walking	44 volunteers walk	City	5 months	GPS	[Shin et al. 2008]
Mobile phone calls	Mobile phone data	City	6 months	Cellular tower	[Gonzalez et al. 2008]
	Mobile phone data	City	3 months	Cellular tower	[Song et al. 2010]
	LA & NYC CDR data	City	3 months	Cellular tower	[Isaacman et al. 2012]
	Cellphone GPS log	City	3 years	GPS	[Fan et al. 2015]
	Yahoo GPS log	City	12 months	GPS	[Shimosaka et al. 2015]
	Nokia MDC data	City	18 months	-	[Zhao et al. 2015b]
	Singapore CDR data	City	2 weeks	Cellular tower	[Jiang et al. 2016]
Social networking	Gowalla Checking-in	-	20 months	GPS	[Cho et al. 2011]
	Brightkite Checking-in	-	30 months	GPS	
	Geo-tagged Twitter	Global	12 months	GPS	[Hawelka et al. 2014]
Taxi	Sweden taxi	City	6 months	GPS	[Jiang et al. 2009]
	Hangzhou taxi	City	12 months	GPS	[Li et al. 2012]
	Beijing taxi	City	-	GPS	[Yuan et al. 2012] [Pan et al. 2013] [Zheng et al. 2013] [Zhao et al. 2015a] [Zhang et al. 2017]
	Rome taxi	City	1 month	GPS	[Zhao et al. 2015a]
	San Francisco taxi	City	1 month	GPS	
	Stockholm taxi	City	1 month	GPS	[Ganti et al. 2013]
	Shanghai taxi	City	1 month	GPS	
	NYC taxi	City	1 day	GPS	[Dimitriou et al. 2016]
Bike	Hangzhou BSS data	City	12 months	Bike station location	[Yang et al. 2016]
	BikeNYC	City	6 months		[Zhang et al. 2017]
Subway	London Oyster card	City	4 months	AFC	[Lathia and Capra 2011]
	Shenzhen subway	City	-	AFC	[Zhang et al. 2015a]
Bus	Lisbon bus data	City	2 months	Bus stop location	[Bhattacharya et al. 2013]
Private Cars	Milano2007	City	1 week	GPS	[Giannotti et al. 2011]
	Pisa2010	City	5 weeks	GPS	
	Octotelematics data	City	12 days	GPS	[Trasarti et al. 2011]
Multisource	Shenzhen CDR data	City	2 months	Cellular tower	[Zhang et al. 2014]
	Shenzhen taxicab	City	24 months	GPS	
	Shenzhen bus	City	12 months	GPS	
	Shenzhen smart card	City	29 months	AFC	
	Beijing CDR data	City	1 year	Cellular tower	[Zhang et al. 2014]
	Beijing taxicab	City	3 years	GPS	
	Beijing bus	City	2 years	GPS	
	Beijing smart card	City	3 years	AFC	

Privacy is an important issue in data related research. To date, privacy preserving methods have been proposed for mobile phone CDR data [Mir et al. 2013; Calabrese et al. 2015]. In general, anonymization is used in the data sets and identity related information would usually be removed when publishing.

5. HUMAN MOBILITY MODELS

As a growing topic, its difficult to formularize an appropriate classification of the mobility models used in the community. In [Asgari et al. 2013], the human mobility studies are divided in to three categories: 1) Trajectory-based; 2) Dynamic proximity networks; and 3) Flow on networks. However, the reason for this categorization is not well illustrated in the paper and its also far from complete and accurate. In [Zhao et al. 2016], a more reasonable classification is presented. The human mobility models are summarized as 1) Synthetic mobility model; 2) Levy walks; 3) The exponential-scaling human mobility model; 4) Gravity model; 5) Radiation model; and 6) Multi-view learning model. While in [Jiang et al. 2016], the models are divided into two categories: trip-based and activity-based. It an impressive categorization but this paper only considered the models from mobile phone data sets. In this survey, the categorization we used is similar to that in [Zhao et al. 2016] with some modifications. Synthetic mobility model is removed because it's not recent study based on data mining. The exponential-scaling human mobility model is removed because it's not a very independent topic. Activity-based model is added to make the survey complete.

5.1. Gravity law

First introduced in 1946 [Zipf 1946], gravity law once was the prevailing method in predicting the human movement between cities. In analogy with Newton's law of gravity, the gravity law assumes that the traffic T_{ij} from locations i to j during certain time is proportional to some power of the population of the source m_i and the destination n_j, and decays with the distance r_{ij} between them as

$$T_{ij} = \frac{m_i^{\alpha} n_j^{\beta}}{f(r_{ij})}$$

where α and β are adjustable exponents and the deterrence function $f(r_{ij})$ is chosen to fit the data [Simini et al. 2012].

Although a bunch of limitations of gravity law model are enumerated in [Simini et al. 2012], such as a lack of theoretical guidance on choosing the deterrence function, gravity law is still shown to be effective to some degree in many applications [Erlander and Stewart 1990; Jung et al. 2008; Garske et al. 2011]. Besides, modifications can be made on gravity law to find the travel pattern in subway system [Goh et al. 2012].

5.2. Radiation model

The radiation model is developed to overcome the limitations of the gravity model [Simini et al. 2012]. The job selection process is referred in building the model. Specifically, the average flux T_{ij} from locations i to j can be computed as:

$$T_{ij} = T_i \frac{m_i n_j}{(m_i + s_{ij})(m_i + n_j + s_{ij})}$$

in which m_i and n_j are the population at i and j respectively. The distance between i and j are denoted as r_{ij}. While s_{ij} stands for the total population in the circle of radius r_{ij} centered at i (excluding the source and destination population).

Given the facts that radiation model is easy to use and can be built without much empirical data, the radiation is not widely used in the studies of human mobility. Researches based on radiation model are far less than that based on gravity model and the following Levy flight model.

5.3. Levy flight

Proposed by the French mathematician Paul Levy, the Levy flight model is to extend the famous Brownian motion to be a more general random walk model. A random walk is a mathematical model of taking successive jumps/flights, each in a random direction. A Levy flight is a particular random walk model that involves two distributions: a uniform distribution for the turning angle (θ_i) and a power-law distribution for the flight length (l_i), that is:

$$P(l_i) \sim l_i^{-\alpha}$$

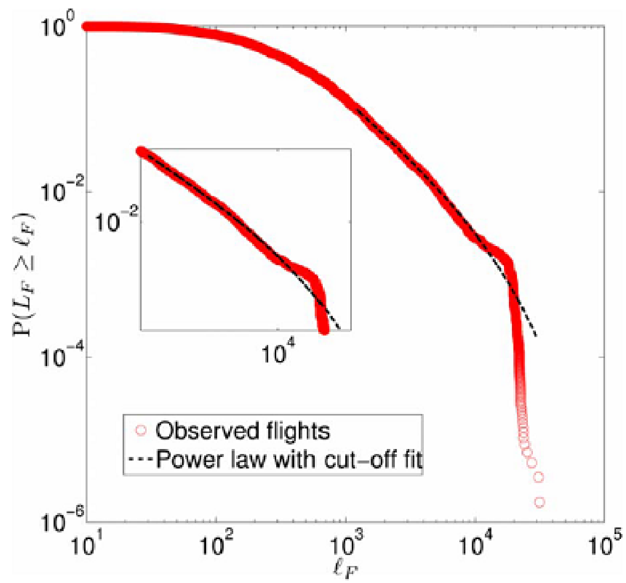

Fig. 1. Observed flights showing a Levy flight length in km behavior [Jiang et al. 2009]

with $1 < \alpha < 3$ [Jiang et al. 2009]. Observation of Levy flight in taxicabs is presented in [Jiang et al. 2009] and Figure 1 shows a typical fitting result.

Levy Flight is a well-studied model to describe the human mobility. Early studies show that Levy flight can be used to model the mobility of animals [Viswanathan et al. 1996; Atkinson et al. 2002; Ramos-Fernández et al. 2004]. While surprisingly, some studies show that human mobility, in an aggregated level, also tend to be Levy flight [Brockmann et al. 2006]. To validate this observation, [Jiang et al. 2009] compared taxicab traces with random walkers, observing that the movement of goal-directed human beings has nothing different compared with random walkers. The reason, according to the author, is the structure of the street network. In [Shin et al. 2008], the author even compared the human mobility with monkeys. They found that some statistical features of human mobility tend to be scale-invariant and self-similar, and a simple Levy walk model can fit the data very well. Besides the structure of the street network, other reasons are also provided to explain the underlying mechanism. In [Han et al. 2011], the author presented a hierarchical geographical model to mimic the real traffic system.

5.4. Multi-view learning

Compared with other models that only rely on the data from a single source, the multi-view learning model tries to incorporate the data from different sources to make the model more accurate [Zhang et al. 2014]. A system called *mPat* was built to fuse the heterogeneous data, extract the human mobility and make real-time prediction based on real-time data supply. Study in some other cities also shows the effectiveness of the model [Zhang et al. 2015b].

Although multi-view data is not a strictly innovative human mobility model, the idea to utilize multi-source data is instructive for the future application and research on human mobility, as more and more heterogeneous data will be collected in future for the goal of smart cities.

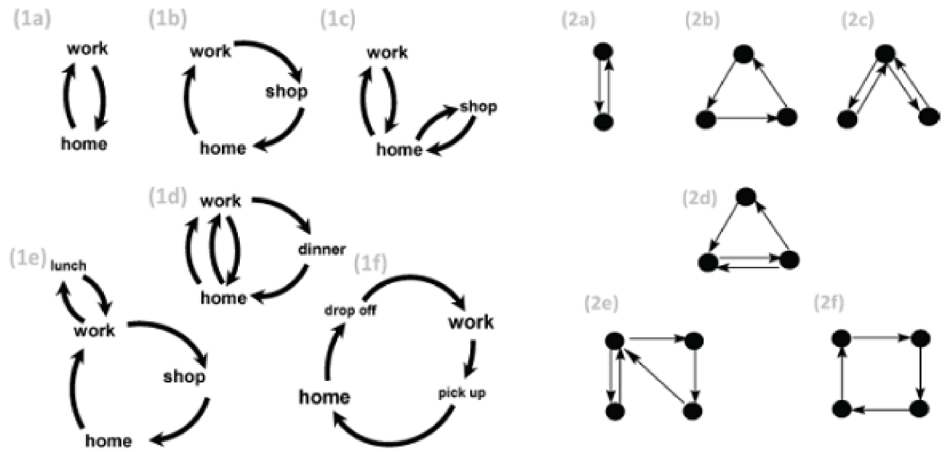

Fig. 2. Examples of daily mobility networks (1a-1f) in daily travel surveys, (2a-2f) in abstract [Jiang et al. 2016]

5.5. Activity based

In [Jiang et al. 2016], the human mobility models are divided into two categories: activity-based and trip-based. For activity-based approach, certain kind of human activities are treated as "motifs", which is closely related to the location one visited. Thus human mobility can be analyzed and classified into different types. Figure 2 illustrates some possible mobility patterns [Jiang et al. 2016].

Although seldom discussed in the computer science literature, activity-based approach is a classical method in transportation area [Pinjari et al. 2011]. Compared with other studies on human mobility, the activity-based methods focus more the reasons that drive people move around. More applications can be developed based on the research with activity-based human mobility patterns.

6. APPLICATION

Unveiling the inherent nature of human mobility can have significant impact on all the phenomenon driven by human mobility. Applications of human mobility modeling can be found in many areas from epidemic control to traffic forecasting, from urban planning to mobile network routing. Besides, human mobility also plays a key role in some novel research area: urban computing and smart cities [Zheng et al. 2014]. Together with the social media, human mobility will be one of the main battlefields in future urban computing applications. In this section, we will review some studies on the application of human mobility.

6.1. Traffic forecasting

Traffic forecasting is the most direct application of the human mobility models. Basically two types of applications can be found in traffic forecasting: to predict the trend of the crowd, and to predict the future movement of a certain individual.

To forecast large group of human is critical in disaster response and urban planning. Large group dynamics in Hajj pilgrimage, the largest annual religious gathering in the world, is analyzed to avoid congestions due to improper arrangements [Jamil et al. 2015]. Similar idea can be found in [Fan et al. 2015], in which a model is built to predict short-term rare behavior such as people gathering for the New Year Eve. Given the tragedy like 2014 Shanghai stampede [Wikipedia 2017], this kind of emergency prediction mechanism will be critical in ensuring the safety for urban residents.

Different models are used in predicting the human mobility. In [Li et al. 2012], a time-series model (ARIMA) is built to forecast the spatial-temporal variation of passengers in a hotspot. Based on the prediction system, taxi drivers can reduce the time taken and distance travelled, to find their next passenger, by 37.1% and 6.4%, respectively. The prediction of taxi demand in urban area is also discussed in [Dimitriou et al. 2016]. Similarly, a Poisson regression model is proposed in [Shimosaka et al. 2015] to deal with the non-linear effects on urban dynamics from external factor. Recently, deep learning is also used to build more complex model to predict the human mobility [Zhang et al. 2017].

Unlike the prediction of group trend, the prediction of individual movement needs a more fine-grained model. In [Zhang et al. 2015a], a fine-grained Bayesian mode is built to predict the destination of a passenger in subway system. Deep learning is also used to predict a person's future movements and their transportation modes in the large-scale transportation network [Song et al. 2016]. Moreover, a real-time framework in predicting human mobility with less volumes of data but produce satisfactory prediction accuracies is proposed in [Kulkarni et al. 2016].

6.2. Urban planning

Mastering the human mobility pattern can help improving urban planning in many ways. The effectiveness of the carried out planning, such as a newly built road and subway lines in a city can be evaluated by monitoring the urban human mobility [Zheng et al. 2011]. Using human mobility pattern, the functions of different regions within the city, such as education and science areas and developing commercial/business/entertainment areas, can be discovered [Yuan et al. 2012]. Similarly, in [Chen et al. 2015], the authors try to find the correlation between the collective human mobility patterns with the function of certain region, which can reduce the possibility of congestion in the urban planning stage. A thorough case study is made in [Jiang et al. 2016] about human mobility pattern in Singapore. Based on the observation in Singapore, planners will be enabled to arrange shopping clusters along certain stops for transit-oriented development, add transportation alternatives and improve level of service along certain corridors, improve community facilities for targeted population.

6.3. Epidemics controlling

To forecast and control the epidemics has long been an important mission for the public health. The studies on this topic usually care about the large group mobility globally. In [Hufnagel et al. 2004], the forecasting of the geographical spread of epidemics is shown to be possible by using a probabilistic model. The endangered regions can be discovered in advance to take control actions. Similar studies can be found in [Colizza et al. 2007], in which prediction is made on human mobility model and compared with some baseline with different containment strategies, including travel restrictions and the therapeutic use of antiviral drugs.

6.4. Other applications

Applications based on human mobility come into a great diversity in recent years and many of them cannot be classified easily to certain categories. Thus here we summarize some of them. With more and more publications come out, it is highly possible that some new categories emerge and we will adjust this section accordingly.

Mobile network routing is one of the precursor application of human mobility. In [Shin et al. 2007], a Levy walk model is built to model the outdoor human mobility. Similar conclusion is also made in [Shin et al. 2008]. In [Zhao et al. 2015a], the routing algorithm varies in regions with different functions, which is discovered by observing the taxi traces.

Carpooling will be an effective solution to the urban congestion if rationally utilized. A carpool strategy is proposed in [Trasarti et al. 2011] by finding and matching the spatially similar trajectories.

In addition to finding the big trend in urban traffic, discovering anomalies sometimes can server for more important purpose. Anomalies in urban traffic anomalies are caused by accidents, control, protests, sport events, celebrations, disasters and other events. In [Pan et al. 2013], anomalies in urban traffic can be detected by monitoring the drivers' routing behavior, together with the data from social media.

Air quality has become a severe problem in some countries such as China and India. In [Zheng et al. 2013], the fine-grained air quality in a city level is estimated by involving the multi-data source. Based on the data, the spatial and temporal of air pollution can also be analyzed.

There is a tendency that crowd sourcing evolving from cellphone based to vehicle based, as Amazon and Uber are conducting crowd delivery. With the predictable human mobility, the performance of crowd sourcing can be improved in future scenarios and new recruiting mechanism can be implemented [He et al. 2015].

Bike sharing system has become an urban infrastructure in many cities, and related studies emerges in recent years from different aspects. In [Yang et al. 2016], the demand in the system is predicted and a rebalancing strategy is proposed.

7. REMAINING ISSUES

Although the community has already made great progress in studies related to human mobility, there are still huge gaps ahead to what we envision of smart cities. Current research on human mobility still tend to be a standalone area and lack connection and cooperation with other sections in a smart city setting. Transportation should be incorporated in urban planning, healthcare and environment management to make more value.

As a new trend in recent years, shared-use mobility such as Uber and Lyft bring great impact on traditional taxi market. Moreover, Amazon and Uber are also working on delivery project which can be a new economic trend. This kind of crowdsourcing can be further studied to meet the practical demand. Thus we need more related researches to shed light on this phenomenon.

REFERENCES

Fereshteh Asgari, Vincent Gauthier, and Monique Becker. 2013. A survey on human mobility and its applications. *arXiv preprint arXiv:1307.0814* (2013).

RPD Atkinson, CJ Rhodes, DW Macdonald, and RM Anderson. 2002. Scale-free dynamics in the movement patterns of jackals. *Oikos* 98, 1 (2002), 134–140.

Vitaly Belik, Theo Geisel, and Dirk Brockmann. 2011. Natural human mobility patterns and spatial spread of infectious diseases. *Physical Review X* 1, 1 (2011), 011001.

Christian Bettstetter, Giovanni Resta, and Paolo Santi. 2003. The node distribution of the random waypoint mobility model for wireless ad hoc networks. *IEEE Transactions on mobile computing* 2, 3 (2003), 257–269.

Sourav Bhattacharya, Santi Phithakkitnukoon, Petteri Nurmi, Arto Klami, Marco Veloso, and Carlos Bento. 2013. Gaussian process-based predictive modeling for bus ridership. In *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication*. ACM, 1189–1198.

Dirk Brockmann, Lars Hufnagel, and Theo Geisel. 2006. The scaling laws of human travel. *Nature* 439, 7075 (2006), 462–465.

Francesco Calabrese, Laura Ferrari, and Vincent D Blondel. 2015. Urban sensing using mobile phone network data: a survey of research. *ACM Computing Surveys (CSUR)* 47, 2 (2015), 25.

Jingmin Chen, Su Yang, Wenshan Wang, and Minjie Wang. 2015. Social context awareness from taxi traces: mining how human mobility patterns are shaped by bags of POI. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*. ACM, 97–100.

Eunjoon Cho, Seth A Myers, and Jure Leskovec. 2011. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1082–1090.

Vittoria Colizza, Alain Barrat, Marc Barthelemy, Alain-Jacques Valleron, and Alessandro Vespignani. 2007. Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions. *PLoS Med* 4, 1 (2007), e13.

Loukas Dimitriou, Christina Christodoulou, Elena Kourti, and Symeon Christodoulou. 2016. Dynamic monitoring of taxi demand profiles, utilizing location-specific information in large metropolitan areas. In *Electrotechnical Conference (MELECON), 2016 18th Mediterranean*. IEEE, 1–6.

Sven Erlander and Neil F Stewart. 1990. *The gravity model in transportation analysis: theory and extensions*. Vol. 3. Vsp.

Zipei Fan, Xuan Song, Ryosuke Shibasaki, and Ryutaro Adachi. 2015. CityMomentum: an online approach for crowd behavior prediction at a citywide level. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 559–569.

Raghu Ganti, Mudhakar Srivatsa, Anand Ranganathan, and Jiawei Han. 2013. Inferring human mobility patterns from taxicab location traces. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*. ACM, 459–468.

Tini Garske, Hongjie Yu, Zhibin Peng, Min Ye, Hang Zhou, Xiaowen Cheng, Jiabing Wu, and Neil Ferguson. 2011. Travel patterns in China. *PloS one* 6, 2 (2011), e16364.

Fosca Giannotti, Mirco Nanni, Dino Pedreschi, Fabio Pinelli, Chiara Renso, Salvatore Rinzivillo, and Roberto Trasarti. 2011. Unveiling the complexity of human mobility by querying and mining massive trajectory data. *The VLDB JournalThe International Journal on Very Large Data Bases* 20, 5 (2011), 695–719.

Segun Goh, Keumsook Lee, Jong Soo Park, and MY Choi. 2012. Modification of the gravity model and application to the metropolitan Seoul subway system. *Physical Review E* 86, 2 (2012), 026102.

Marta C Gonzalez, Cesar A Hidalgo, and Albert-Laszlo Barabasi. 2008. Understanding individual human mobility patterns. *Nature* 453, 7196 (2008), 779–782.

Xiao-Pu Han, Qiang Hao, Bing-Hong Wang, and Tao Zhou. 2011. Origin of the scaling law in human mobility: Hierarchy of traffic systems. *Physical Review E* 83, 3 (2011), 036117.

Bartosz Hawelka, Izabela Sitko, Euro Beinat, Stanislav Sobolevsky, Pavlos Kazakopoulos, and Carlo Ratti. 2014. Geo-located Twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science* 41, 3 (2014), 260–271.

Zongjian He, Jiannong Cao, and Xuefeng Liu. 2015. High quality participant recruitment in vehicle-based crowdsourcing using predictable mobility. In *Computer Communications (INFOCOM), 2015 IEEE Conference on*. IEEE, 2542–2550.

Andrea Hess, Karin Anna Hummel, Wilfried N Gansterer, and Günter Haring. 2016. Data-driven human mobility modeling: a survey and engineering guidance for mobile networking. *ACM Computing Surveys (CSUR)* 48, 3 (2016), 38.

Lars Hufnagel, Dirk Brockmann, and Theo Geisel. 2004. Forecast and control of epidemics in a globalized world. *Proceedings of the National Academy of Sciences of the United States of America* 101, 42 (2004), 15124–15129.

Sibren Isaacman, Richard Becker, Ramón Cáceres, Margaret Martonosi, James Rowland, Alexander Varshavsky, and Walter Willinger. 2012. Human mobility modeling at metropolitan scales. In *Proceedings of the 10th international conference on Mobile systems, applications, and services*. ACM, 239–252.

Shuja Jamil, Anas Basalamah, Ahmed Lbath, and Moustafa Youssef. 2015. Hybrid participatory sensing for analyzing group dynamics in the largest annual religious gathering. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 547–558.

Bin Jiang, Junjun Yin, and Sijian Zhao. 2009. Characterizing the human mobility pattern in a large street network. *Physical Review E* 80, 2 (2009), 021136.

Shan Jiang, Joseph Ferreira, and Marta C Gonzales. 2016. Activity-based human mobility patterns inferred from mobile phone data: A case study of Singapore. *IEEE Transactions on Big Data* (2016).

Woo-Sung Jung, Fengzhong Wang, and H Eugene Stanley. 2008. Gravity model in the Korean highway. *EPL (Europhysics Letters)* 81, 4 (2008), 48005.

Vaibhav Kulkarni, Arielle Moro, and Benoît Garbinato. 2016. A mobility prediction system leveraging realtime location data streams: poster. In *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*. ACM, 430–432.

Neal Lathia and Licia Capra. 2011. How smart is your smartcard?: measuring travel behaviours, perceptions, and incentives. In *Proceedings of the 13th international conference on Ubiquitous computing*. ACM, 291–300.

Xiaolong Li, Gang Pan, Zhaohui Wu, Guande Qi, Shijian Li, Daqing Zhang, Wangsheng Zhang, and Zonghui Wang. 2012. Prediction of urban human mobility using large-scale taxi traces and its applications. *Frontiers of Computer Science* 6, 1 (2012), 111–121.

Darakhshan J Mir, Sibren Isaacman, Ramón Cáceres, Margaret Martonosi, and Rebecca N Wright. 2013. Dp-where: Differentially private modeling of human mobility. In *Big Data, 2013 IEEE International Conference on*. IEEE, 580–588.

Bei Pan, Yu Zheng, David Wilkie, and Cyrus Shahabi. 2013. Crowd sensing of traffic anomalies based on human mobility and social media. In *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 344–353.

Abdul Rawoof Pinjari, Chandra R Bhat, and others. 2011. Activity-based travel demand analysis. *A Handbook of Transport Economics* 10 (2011), 213–248.

Gabriel Ramos-Fernández, José L Mateos, Octavio Miramontes, Germinal Cocho, Hernán Larralde, and Barbara Ayala-Orozco. 2004. Lévy walk patterns in the foraging movements of spider monkeys (Ateles geoffroyi). *Behavioral Ecology and Sociobiology* 55, 3 (2004), 223–230.

Masamichi Shimosaka, Keisuke Maeda, Takeshi Tsukiji, and Kota Tsubouchi. 2015. Forecasting urban dynamics with mobility logs by bilinear Poisson regression. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 535–546.

Injong Rhee NCSU Minsu Shin, Seongik Hong NCSU Kyunghan Lee, and Song Chong. 2007. Human mobility patterns and their impact on routing in human-driven mobile networks. In *HotNets-VI*.

RM Shin, S Hong, K Lee, and S Chong. 2008. On the Levy-walk nature of human mobility: Do humans walk like monkeys?. In *Proc. IEEE INFOCOM*. 924–932.

Filippo Simini, Marta C González, Amos Maritan, and Albert-László Barabási. 2012. A universal model for mobility and migration patterns. *Nature* 484, 7392 (2012), 96–100.

Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. 2010. Limits of predictability in human mobility. *Science* 327, 5968 (2010), 1018–1021.

Xuan Song, Hiroshi Kanasugi, and Ryosuke Shibasaki. 2016. Deeptransport: Prediction and simulation of human mobility and transportation mode at a citywide level. IJCAI.

Roberto Trasarti, Fabio Pinelli, Mirco Nanni, and Fosca Giannotti. 2011. Mining mobility user profiles for car pooling. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1190–1198.

Gandhimohan M Viswanathan, V Afanasyev, SV Buldyrev, EJ Murphy, and others. 1996. Lévy flight search patterns of wandering albatrosses. *Nature* 381, 6581 (1996), 413.

Wikipedia. 2017. 2014 Shanghai stampede — Wikipedia, The Free Encyclopedia. (2017). https://en.wikipedia.org/w/index.php?title=2014_Shanghai_stampede&oldid=757788218 [Online; accessed 1-January-2017].

Zidong Yang, Ji Hu, Yuanchao Shu, Peng Cheng, Jiming Chen, and Thomas Moscibroda. 2016. Mobility Modeling and Prediction in Bike-Sharing Systems. In *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 165–178.

Jing Yuan, Yu Zheng, and Xing Xie. 2012. Discovering regions of different functions in a city using human mobility and POIs. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 186–194.

Desheng Zhang, Jun Huang, Ye Li, Fan Zhang, Chengzhong Xu, and Tian He. 2014. Exploring human mobility with multi-source data at extremely large metropolitan scales. In *Proceedings of the 20th annual international conference on Mobile computing and networking*. ACM, 201–212.

Desheng Zhang, Riiobing Jiang, Shiiai Wang, Yanmin Zhu, Bo Yang, Jian Cao, Fan Zhang, and Tian He. 2015a. EveryoneCounts: Data-driven digital advertising with uncertain demand model in metro networks. In *Big Data (Big Data), 2015 IEEE International Conference on*. IEEE, 898–907.

Desheng Zhang, Juanjuan Zhao, Fan Zhang, and Tian He. 2015b. comobile: Real-time human mobility modeling at urban scale using multi-view learning. In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 40.

Junbo Zhang, Yu Zheng, Dekang Qi, Ruiyuan Li, Xiuwen Yi, and Tianrui Li. 2017. Predicting Citywide Crowd Flows Using Deep Spatio-Temporal Residual Networks. *arXiv preprint arXiv:1701.02543* (2017).

Kai Zhao, Mohan Prasath Chinnasamy, and Sasu Tarkoma. 2015a. Automatic city region analysis for urban routing. In *Data Mining Workshop (ICDMW), 2015 IEEE International Conference on*. IEEE, 1136–1142.

Kai Zhao, Mirco Musolesi, Pan Hui, Weixiong Rao, and Sasu Tarkoma. 2015b. Explaining the power-law distribution of human mobility through transportation modality decomposition. *Scientific reports* (2015).

Kai Zhao, Sasu Tarkoma, Siyuan Liu, and Huy Vo. 2016. Urban Human Mobility Data Mining: An Overview. In *2016 IEEE International Conference on Big Data (Big Data)*. IEEE.

Yu Zheng. 2015. Methodologies for cross-domain data fusion: An overview. *IEEE transactions on big data* 1, 1 (2015), 16–34.

Yu Zheng, Licia Capra, Ouri Wolfson, and Hai Yang. 2014. Urban computing: concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)* 5, 3 (2014), 38.

Yu Zheng, Furui Liu, and Hsun-Ping Hsieh. 2013. U-air: When urban air quality inference meets big data. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1436–1444.

Yu Zheng, Yanchi Liu, Jing Yuan, and Xing Xie. 2011. Urban computing with taxicabs. In *Proceedings of the 13th international conference on Ubiquitous computing*. ACM, 89–98.

Yu Zheng, Lizhu Zhang, Xing Xie, and Wei-Ying Ma. 2009. Mining interesting locations and travel sequences from GPS trajectories. In *Proceedings of the 18th international conference on World wide web*. ACM, 791–800.

George Kingsley Zipf. 1946. The P 1 P 2/D hypothesis: on the intercity movement of persons. *American sociological review* 11, 6 (1946), 677–686.