

ON THE CHALLENGES OF MOBILITY PREDICTION IN SMART CITIES

H. Boukhedouma¹, A. Meziane¹, S. Hammoudi², A. Benna¹

¹CERIST, Algiers, Algeria - (hboukhedouma, ameziane, abenna)@cerist.dz

²ESEO-TECH - ERIS TEAM, Angers, France, slimane.hammoudi@eseo.fr

Commission IV

KEY WORDS: movement type, smart city, context, prediction, mobility

ABSTRACT

The mass of data generated from people's mobility in smart cities is constantly increasing, thus making a new business for large companies. These data are often used for mobility prediction in order to improve services or even systems such as the development of location-based services, personalized recommendation systems, and mobile communication systems. In this paper, we identify the mobility prediction issues and challenges serving as guideline for researchers and developers in mobility prediction. To this end, we first identify the key concepts and classifications related to mobility prediction. We then, focus on challenges in mobility prediction from a deep literature study. These classifications and challenges are for serving further understanding, development and enhancement of the mobility prediction vision.

1. INTRODUCTION

In the recent past, the appearance of smart cities and internet of things (IoT) systems along with new technologies (e.g., mobile networks-*MN*, sensor networks), and new tools (e.g., smartphone), has led to an impressive growth of amount data and information produced. This amount of data tends to multiply wildly in the case of smart cities which include at the same time these new concepts, technologies and tools. Indeed, a smart city features the utilization of information and communication technology (ICT) infrastructure, human resources, social capital, and environment resources for economic development, and high quality of human life. Thus, analysis and mining of sensed data from dynamic cities is an important step towards making a city smart (Pan et al., 2013). Among the flowing data in such a city, those linked to individual movements (mobility data) are very interesting for a large community of researchers and developers, especially in the mobility prediction field.

Predicting a mobile user location is an inherently interesting and challenging problem in several domains such as the development of location-based services, personalized recommendations, suspicious target tracking, intelligent transportation and mobile communication systems (MCSs). For example, in MCSs, location prediction has received increased attention driven by applications in location management, call admission control, smooth handoffs, and resource reservation for improved quality of service (Samaan, Karmouch, 2005).

However, predicting mobility requires the availability of a large amount of data from very heterogeneous sources, especially when it comes to a smart city. Two levels of collection can be distinguished in the data storage location. The first level concerns the data acquisition by a system or by an application from a mobile device, such in the case of the Mobile Crowd Sensing and Computing (MCSC) paradigm or even with the use of a recommendation application on a smartphone. In this case, the data are stored in the mobile device or in an external storage place related to the application. The second level is related to a collection of data from a storage device used by a system, such

as MCSs, smart cards management systems, etc. In this case, the data, in particular mobility data, are stored in specific equipment (e.g., HLR¹, VLR², sensor nodes, etc.) and are collected directly from this equipment. However, because of the privacy concerns of this type of data, they may be subject to constraints and conditions when accessing to those data. A major challenge is to access and recover the. In addition, data are often stored in a raw state such as log files (Zheng et al., 2010). So, before being exploited, data stored in log files must be transformed into other formats like GPX or PLT format.

Once data is collected and stored, prediction requires a model coherent with the prediction's application domain and able to provide the best mobility prediction in terms of accuracy, cost, etc. A prediction model can be produced based on one of the usual techniques dedicated to prediction such as Markov chains-*MCs* (Amirrudin et al., 2013a; Qiao et al., 2015), Machine Learning-*ML* (Ozturka et al., 2019), Bayesian Networks-*BNs* (Dash et al., 2015), and data mining-*DM* techniques (Mcinerney et al., 2013), where most of them are based on learning from previous data to predict user mobility.

Although the mobility prediction has been the subject of several research works which sometimes gave acceptable results, in terms of precision (Amirrudin et al., 2013b;), and cost, certain issues remain open and accept new contributions.

In this paper, we aim to focus on the main concepts related to the mobility, data required for mobility prediction and on related works on mobility prediction. We also aim to disclose issues allowing researchers and developers to orient themselves towards open questions on the mobility prediction.

The remainder of this paper is organized as follows: Section 2 presents the basics of mobility prediction. Section 3 gives an overview on the mobility prediction works. In section 4 we

¹http://fr.wikipedia.org/w/index.php?title=Home_Location_Register&oldid=162312017

²http://fr.wikipedia.org/w/index.php?title=Visitor_Location_Register&oldid=105701924

propose a classification for mobility prediction. In Section 5 we focus on a set of challenges related to mobility prediction. And in Section 6 we give a conclusion and some perspectives.

2. DEFINITIONS AND KEY CONCEPTS

In this section we introduce the key points of mobility prediction. We start by a definition and a classification of mobility. Then we focus on the data required for mobility prediction.

2.1 Mobility

In general, the definition of mobility can be obtained from a dictionary or an encyclopedia. In *Larousse*³ editions, for example, the definition can be translated as follows: "A Character of what is susceptible to movement, of what can move or be moved, change place, function."

In the context of communication networks, we carry over the definition given by *Samuel Pierre* in his book (Samuel, 2003): "In the domain of communication networks, mobility can be defined as the ability to access, from any place, all the services normally available in a fixed and wired environment such as a home or an office. These services include, among other things, the possibility of conducting a telephone conversation while driving a car, being reached from a traditional telephone or an IP address anywhere in the world, or receiving e-mail, faxes or voice messages while traveling abroad."

From our viewpoint, in the context of a smart city, mobility can be defined as any movement an entity undergoes over time, where an entity can be an object or an individual (a person). This movement can vary from a simple action (move the hand, take a seat, get up...) to a real trip (walk from a point A to point B). In reality, in such a case, mobility is a relative notion in the sense that it is closely linked to the size of the environment in which we want to define it. For example, if the environment consists of a city, the mobility is considered as being the movements from a departure point to an arrival point. However, when the environment is restricted to a limited space (house, bedroom...), the mobility concerns elementary actions taking place in this space. The movements can also be repeated over time either in the same way or in different ways: we therefore speak about random or regular movements.

2.2 Mobility classification

Based on the size of the environment in which the movement takes place, on the way in which the movement is carried out and on works dealing with movement, we distinguish two types of mobility classification: according to the distance travelled and according to the nature of the movement (see Figure 1).

Classification according to the distance travelled (environment size): based on the environment size, we distinguish: *extended* mobility and *restricted* mobility. *Extended* mobility is characterized by movements spanning a long distance such as the movement of an individual from his home to his workplace. *Restricted* mobility is characterized by movements taking place in a limited area (an office, a bedroom, etc.). Most of the research works dealing with mobility, in particular mobility prediction (Jiang et al., 2016; Liou, Huang, 2005) are directed towards *extended* mobility. However, some

works which are interested in *restricted* mobility exist (Almeida, Azkune, 2018). In a smart city, *restricted* mobility can be related to connected objects (IoT).

Classification according to the nature of the movement: refers to the way in which the movement recurs. Here, we distinguish *regular* mobility and *random* mobility. *Regular* mobility concerns movements reproducing in the same way (e.g. taking the same path to get to work). In contrast, *random* mobility concerns either new movements or movements that do not reproduce in the same way (for e.g., taking two different paths to get home). The number of works directed towards regular mobility (Nadembega et al., 2014) is greater compared to the number of works dealing with random mobility (Liu et al., 1998), or even compared to those treating both types of mobility (Liu, Maguire, 1996). Also, for mobility prediction, the accuracy is better in the case of regular mobility (Jiang et al., 2016). Random mobility is, therefore, a real challenge in the area of mobility prediction.

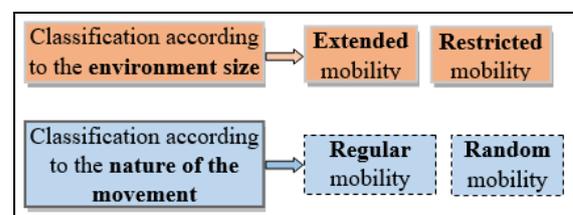


Figure 1. Mobility classification

2.3 Mobility prediction data

Basic prediction models rely on historical mobility data to provide predictions (Anisetti et al., 2011). Other models rely on contextual data, in addition to historical mobility data (Abu-Ghazaleh, Alfa, 2009). Another category includes models that only use contextual data (Samaan, Karmouch, 2005). However, little works has been done in this latter category.

In this section, we aim to discover provenance, storage and exploitation of mobility data. Also, we define contextual data, and give an overview about their use for mobility prediction.

2.3.1 Mobility data: provenance, storage and exploitation

The study of any mobile object or individual's mobility requires mainly a set of data related to its location at different moments. In a smart city, the mass of mobility data circulating in the city is constantly increasing. However, before being used, these data, coming from different sources (mobile devices, vehicles, magnetic cards, sensors), must be collected and processed using specific tools, techniques and technologies (GPS, RFID, etc.).

Data provenance

In (Pan et al., 2013), the main data sources have been grouped into four categories, namely mobile devices (smartphones, laptop...), vehicles equipped with global positioning system (GPS) devices, smart cards (bank cards and transport cards), and floating sensors. Other sources exist in a smart city, such as cameras installed in the city, providing a big quantity of videos. Acquiring data from these different sources differs depending on the source. The most used techniques and technologies are GPS, WiFi, GSM and radio frequency identification (RFID). Based on (Pan et al., 2013), we summarize in Table 1 the main

³<https://www.larousse.fr/dictionnaires/francais/mobilité/51890>

data sources as well as the data collection techniques used for each type of sources.

Data source	Location method
Mobile devices	GPS, WiFi, GSM, Bluetooth
Vehicles equipped with GPS	GPS
Smart cards	Location of the machine <i>e.g. Automated Teller Machine</i>
Floating sensors	RFID

Table 1. Data source and relating collection methods

Data storage

Data can be often collected from specific equipment constituting part of the systems architecture to which the different data sources are linked. For example, mobile devices are necessarily linked to a mobile network, and their locations data are stored in specific equipments (HLR, VLR, etc.). Also, in the case of sensors, belonging to a specific sensor network, the location is stored in the sensors themselves, in the base stations or even in the processing centers. Because of the huge amount of data, Edge computing solution and Artificial Intelligence techniques are to be considered.

Data exploitation

Collected data are initially in a raw state. In order to use them, they must be transformed and saved in an exploitable format. For example, in the *GeoLife* project (Zheng et al., 2010), mobility data related to individual's trajectories are generated in the form of *GPS logs* and are transformed and saved in *PLT* files. The main fields recorded in a *GPS log*⁴ file are: longitude, latitude, altitude, current date and time of day. The data in the *GPS log* files are saved in *NMEA*⁵ format. To use them, they must be converted to other formats such as *GPX* or *PLT* format using tools, software or even websites, such as the *Logcat* utility, the *GPSTabel* software or the *GPS Visualizer* website. Also, to exploit data, consideration of data models (e.g. Fiware data model) is very important in the context of a smart city. Figure 2 illustrates the steps for producing usable data.

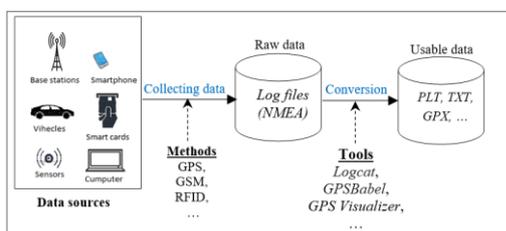


Figure 2. Process of usable data production in smart city

2.3.2 Contextual data

The context has been defined in several ways. The definition most often cited in the literature is that given in (Abowd et al., 1999): « *Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, object that is considered relevant to the interaction*

between a user and an application, including the user and applications themselves »

Zimmermann et al (2007), propose a new definition specifying five categories of contextual information. Moreover, Dey's definition is proposed in the context of interaction between a user and an application. However, nowadays with smartphones, the system is able to give relevant information to the user without explicit interaction. According to Zimmermann, « *Context is any information that can be used to characterize the situation of an entity. Elements of description of this context information fall into five categories: individuality, activity, location, time, and relations* ».

Figure 3 depicts an entity as well as its different contextual categories according to the vision of Zimmermann et al.

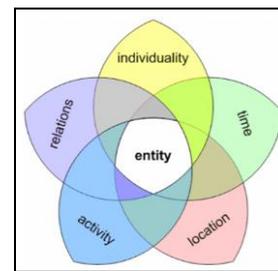


Figure 3. The five basic categories for context information (Zimmermann et al., 2007)

When it's about mobility, the context concerns any information susceptible to inform about individual's movements or which influences their movements. In this perspective, the individual's movements depend on several parameters such as profile (age, profession, preferences, etc.), time, location, environment (weather information, information on road traffic, etc.), or again means of travel (on foot, by car, public transport, etc.). For example, to get to work, a person can take different paths under different conditions (rain, obstacle, etc.). Thus, to predict the individual's movements, considering these conditions is necessary in order to improve the prediction accuracy.

Contextual information can be obtained from a variety of sources. Benouaret (2017) distinguishes three types of contextual information: explicit, implicit and inferred. For explicit sources, context information is already included in the data or directly requested from the user. The most obvious example here corresponds to a user registration on a system, which provides personal information. For implicit sources, information is obtained from the data or the environment in which a user is situated without explicitly asking him for this information. For example, we can get the geographic situation of a user using an application installed on their smartphone. Concerning inferred sources, information is obtained using data exploitation and exploration methods such as data mining techniques.

3. OVERVIEW OF THE RELATED WORKS

The establishment of a prediction model or algorithm represents a key point for mobility prediction. In this section, we summarize in Table 2 some mobility prediction works carried out based on Markov models (standard and hidden MCs), BNs (standard and dynamic) and ML techniques (Artificial Neural Networks *ANN*, Deep learning *DL*, Recurrent Neural Networks *RNN*). We give also the main conclusions obtained in the cited works.

⁴<https://desktop.arcgis.com/fr/arcmap/10.3/map/display-gps-data/gps-information-in-logs.htm>

⁵<https://www.dstar-france.fr/randonnees/gps-convertir-fichier-log>

Table 2 is established on the basis of a study described in (Zhang, Dai, 2018), summarizing some works dealing with mobility prediction. The following criteria are considered: objective, technique used, movement type, context consideration, precision, and complexity/costs generated (calculation time, memory space...). The "technique used" and "precision" criteria are reported from (Zhang, Dai, 2018). The remaining criteria are new and are useful to the identification of mobility prediction issues.

From our study of the works cited in Table 2, we retain the following:

- a- All works are oriented towards the application of mobility prediction to mobile networks (except Qiao et al., 2015).
- b- Markov models (standard and hidden) are widely used in the field of mobility prediction. Standard MCs are simple and easy to implement (Zhang, Dai, 2018) but their performance, in terms of precision, are often subject to certain constraints (transition matrix's values (Amirrudin et al., 2013a), movement type (Jiang et al., 2016) ...). The hidden Markov models are also efficient in terms of accuracy (Zhang, Dai, 2018) (about 53% in (Lv et al., 2014), sometimes exceeds 80% in (Qiao et al., 2015) and even greater than 90% in (Amirrudin et al., 2013b)). However, their complexity can increase with the increase of the number of hidden states and the size of the history (Lv et al., 2014), like in Ultra-Dense mobile Networks (UDNs), because of the complexity of transition matrix which considers hidden and observable states (Zhang, Dai, 2018).
- c- Standard or hidden Markov models-based approaches often use a *Clustering* to determine the regions of interest. In (Gambis et al., 2012) and (Jiang et al., 2016) authors used DJCluster Algorithm; in (Qiao et al., 2015) authors used a clustering analogous to DBSCAN (Zhang et al., 2009).
- d- Bayesian networks gave good results in terms of prediction accuracy, which has sometimes exceeded 75% (Dash et al., 2015). However, with the dense deployment of small cells (UDNs), the cell environment would be more complex, which would make it more difficult to build a BN (Zhang, Dai, 2018).
- e- Neural networks are known by well-studied algorithms and are famous for their adaptive and self-organization characteristics. However, their algorithms are also known for their amazing computational complexity, especially in case of many hidden layers, which require a lot of learning time to adjust the weight of neurons. In addition, when using ANN in the field of MCSs (in particular UDNs), the procedures of acquiring user positions are influenced by the deployment of this type of network, knowing that position is a paramount parameter for ANN in case of mobility prediction (Zhang, Dai, 2018). Certain works based on the NNs provided acceptable results such as the model proposed in (Liou, Huang, 2005) and (Wickramasuriya et al., 2017) in which the precision exceeded 95% and reached 98% respectively. In contrast, the results provided by other works are not very satisfactory compared to the costs generated by such a technique, like the model proposed in (Parija et al., 2013) that allows to provide results only for regular movements.
- f- Deep learning is one of the ML models based on in-depth learning. Currently, DL is present in several fields, notably

in the medical sector (image processing and classification), in the computer-assisted surveillance field (facial recognition) and in mobility prediction (Ozturka et al., 2019). In turn, the models based on DL have also shown acceptable results such as in (Ozturka et al., 2019).

In the next section we aim to propose a classification of mobility prediction according to our study (state of the art) on the works dealing with mobility prediction.

4. MOBILITY PREDICTION CLASSIFICATION

With a lack of a reference classification of prediction models and standard criteria on which the classification of mobility prediction is based, we propose a classification based solely on existing prediction work. For that, we define the following four criteria: the use of history, the technique used, the use of deduction rules and the data type used. According to these criteria, we propose the following classifications (Figure 4):

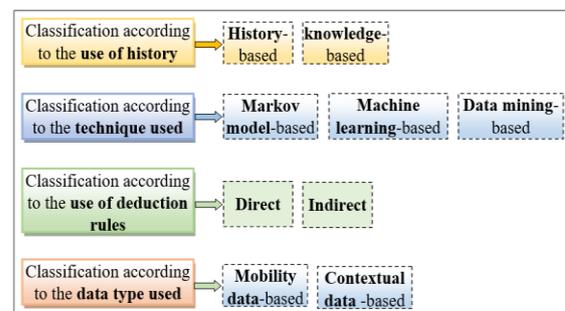


Figure 4. Mobility prediction classification

Classification according to the use of history: we distinguish historical-based models (Boc et al., 2011; Abu-Ghazaleh, Alfa, 2009), and knowledge-based models (Samaan, Karmouch, 2005). Historical-based models are models that use the historical data for predicting mobility. Knowledge-based models use other data such future planning, future goals, etc. to predict mobility.

Classification according to the technique used: works are categorized according to the technique used. We distinguish, among others, Markov models-based works (standard and hidden Markov chains) (Amirrudin et al., 2013b; Lv et al., 2014), Bayesian networks-based works (Dash et al., 2015), machines learning-based works (Ozturka et al., 2019), data mining based works (Soh et al., 2006), etc.

Classification according to the use of deduction rules: we distinguish direct prediction models and indirect prediction models. Direct prediction models are models that use the data of the person concerned by prediction. Most of the works cited in this paper are direct prediction-oriented. Indirect prediction models are models whose prediction is based on deduction rules, such as profile similarity (Chamek et al., 2012).

Classification according to the data type used: we distinguish mobility data-based models (Anagnostopoulos et al., 2012), and contextual data-based models (Göndör et al., 2013). Mobility data-based models are those using only mobility data in order to predict mobility. Contextual data-based models are models whose prediction takes into account other data, qualified as contextual data (such as, environment information and means of travel), in addition to mobility data.

Works (ref)	Objective	Technique used	Movement type	Context	Precision	Costs/ Complexity	Other
Amirrudin et al., 2013a	Handover management.	Standard MCs	NM	No	Depends on assumed values of the transition matrix	NM	Easy to develop
Amirrudin et al., 2013b	Handover management.	Standard MCs	Regular and random	No	Regular movement: very high (> 95%). Random movement: low (<30%).	NM	Easy to develop
Gambs et al., 2012	Performance improvement	n-MMC : n order Mobility Markov chains	NM	No	n ≤ 2 : high (70% - 95%). otherwise : low	n>2 : Complex calculation. Significant memory usage.	Satisfactory results for n≤2. Precision does not increase too much for n>2. Use of <i>Clustering</i>
Jiang et al., 2016	Location-based services	n-MMC	Regular and random	Standard : Position, date, time.	n=2 and regular movement : well	NM	Use of Clustering
Lv et al., 2014	Performance improvement in LTE network	Hidden MCs	Regular	No	Acceptable (53%)	The complexity increases with the increase in the number of hidden states and the size of the history	Accuracy May Increase in a Real Communication System
Qiao et al., 2015	Transportation network Services improvement	Hidden MCs	NM	Standard : Speed	Acceptable	High execution time	Use of Clustering
Parija et al., 2013	Performance improvement	Standard ANNs	Regular	Standard : distance traveled, direction	Regular movement: satisfactory result	satisfactory result	Model not applicable to random movement.
Liou, Huang, 2005	Improvement of bandwidth performance.	Standard ANNs	NM	Standard : Position, speed, acceleration, direction	Highway: very high (>95%).	Non-complex calculation	High precision for regular movement such as in highway
Wickramasuriya et al., 2017	Handover management and Performance improvement in 5G network	Recurrent NNs	NM	No	Very high (98%): under the assumption that a mobile node can only connect to the three closest BSs.	Complex calculation at the start of learning.	
Ozturka et al., 2019	Handover management.	Recurrent DL	NM	Standard : Coordinates in a grid, date, time	High: under the assumption that the number of neurons in the hidden layers is high	Costs increase as the value of the user dissatisfaction parameter increases	Ideal case: number of neurons in hidden layers = 64.
Dash et al., 2015	Performance improvement	Dynamic BNs	Regular and random	Standard : Position, day of week, time of day	Regular movement and historical size ≤ 2: high (>75%)	Yes: for a dense network	
Akoush, Sameh, 2007	Reduction of the network management costs (handover, ...)	Bns+ANNs (hybrid)	NM	No	Best: relative to other works	NM	

Table 2. Summary on mobility prediction work

In Table 2, MC means *Markov chain*, ANN means *Artificial neural network*, BN means *Bayesian network*, DL means *Deep learning*, NM means *Not Mentioned*, i.e. information about the criteria is not mentioned in the associated work.

In the next section, we focus on the challenges of mobility prediction by identifying six issues obtained from our analysis of works dealing with mobility prediction.

5. MOBILITY PREDICTION CHALLENGES IN SMART CITIES

In this section, we aim to present open questions, related to the mobility prediction, which we considered important according to our study on the mobility prediction. These challenges either concern issues that are already treated but not resolved correctly or those which are not treated yet. Also, these challenges are organized based on trajectory size (all trajectory or a segment only), movement type, context consideration, application domains, evaluation of models and privacy of users data.

5.1 Challenge 1 - Predicting a trajectory or a segment of a trajectory

Prediction of a single transition, a sequence of a trajectory or even an entire trajectory of a mobile individual depends mainly on the use of historical data of their mobility. Indeed, some works (Anisetti et al., 2011) have opted for the use of these data, while others (Lytrivis et al., 2011) have not based on these historical data.

Historical-based works are precise but suffers from high overhead costs linked to constant monitoring requiring a more detailed analysis of the history using data mining and knowledge discovery techniques.

Non historical-based works allow to predict only the final destination and not the trajectory towards this destination (Nadembega et al., 2014). These works don't require historical mobility data, but requires other essential data, such as contextual data, and can also generate very high costs linked to an immense amount of data which must be collected and processed (e.g. preferences, final objective, and planning).

Also, other works (Anagnostopoulos et al., 2012) have proposed models that consider the use of historical mobility data and current network conditions (this is a kind of hybridization between the two types mentioned above). These models suffer from some limitations. They offer either short-term predictions (one transition or two at most), or a long-term prediction with very significant additional costs linked to historical data and processing.

Thus, the challenge here is to be able to propose a long-term prediction model at the best cost (which ensures the best ratio: long-term prediction/costs incurred).

5.2 Challenge 2 – Movement type

Exploring the literature on mobility prediction works allowed us to conclude that most of works performed on the mobility prediction consider regular movements of individuals. In (Samaan, Karmouch, 2005) the authors claim that most of the existing approaches assume that the user travels according to a previously known pattern with regularity. Works based on the assumption of regularity of movement offers acceptable results (between 50% and 70% precision) (Jiang et al., 2016; Lv et al., 2014...), even very satisfactory sometimes (up to 90% accuracy) (Amirrudin et al., 2013b; Liou, Huang, 2005), in terms of prediction accuracy. These works were tested even with random movement. According to authors of these models, prediction accuracy decreases when the regularity of the movement decreases. In other words, these models provide poor results when the movement is random.

Although the assumption of regularity of movement is valid in many cases due to people's daily routine, the assumption that an individual's movements may be random should not be excluded. For example, a tourist who visits a country for the first time can take several different paths, especially during the first days after his arrival. Certain works, such as (Liu et al., 1998) consider the random movement. They proposed models answering this type of movement. To the best of our knowledge, the number of these works remains negligible compared to the models dealing with regular movement. Also, by analyzing the results of the last cited work, we note that the prediction precision is not satisfactory, although the authors claim to have obtained good results.

Another category of works, like (Liu, Maguire, 1996), consider both regular and random movement at the same time. The results are acceptable only for regular movements.

Thus, a major challenge is to be able to propose a model making it possible to predict mobility of a user who habitually moves in a random manner (tourist, transporter, etc.).

5.3 Challenge 3 – Context-based prediction

Most of the proposed works in the field of mobility prediction such (Ozturka et al., 2019; Wang et al., 2019) do not take into account contextual information, or, in the best of cases, consider only some standard information, such as day of the week (work day or weekend day), date, time, speed and profile

(preferences, goals,...) (Dash et al., 2015; Chamek et al., 2012; Du et al., 2011). For example, in (Chamek et al., 2012), the profile is considered, in (Dash et al., 2015), only time is considered, in (Du et al., 2011), time (hour of the day) and day of the week are considered. The second category of works gave better results in terms of precision.

Besides this standard information, there is other information related to more important parameters which directly influence the individual's movements but which, to our knowledge and so far, are not taken into account, except for the work described in (Göndör et al., 2013) which considers meteorological information. The main parameters having a direct impact on the movements of individuals relate to environmental (meteorological) information, means of travel and road traffic. If we take the example of the individual who goes to work, in rainy weather, probably this individual will take a path different from his habitual path (a shortcut) to get to work. Also, if he moves by walking, he can take a different path than when he moves by car or by bicycle.

From the previous example taken from our daily life, it is very clear that contextual information play an important role in predicting mobility and that certain information are more important than others. To reinforce this view point, certain recent works, such as (Wang et al., 2019), dealing with the problem of mobility prediction envisage in their perspectives to take into account contextual information, in particular meteorological information.

Thus, the challenge here consists in proposing a solution which considers contextual information, in particular meteorological information (weather), information related to the means of travel and those related to road traffic. However, collection of this type of data is in itself another challenge. In addition, it is very difficult to consider all of these types of contextual information at the same time.

5.4 Challenge 4 – Application domains

Several applications of mobility prediction have been cited in the literature. According to (Wang et al., 2019), prediction can be applied to personalized recommendations, suspicious target tracking and intelligent transportation. Prediction also plays a big role in the field of mobile networks (Samaan, Karmouch, 2005; Liu et al., 1998...). According to (Samaan, Karmouch, 2005), the importance of mobility prediction techniques can be seen both at the network level (handoff management, resource allocation...) and service level (pushed online advertising, mapping/route guidance...). Also, in (Gambs et al., 2012), the authors proposed an extended prediction model having several potential applications such as the evaluation of geo-privacy mechanisms, the development of location-based services anticipating the next movement of a user and the design of location-aware proactive resource migration.

Although the fields of mobility prediction application are numerous, to the best of our knowledge, all the works have been directed towards the application of mobility prediction in the field of MCSs, such in mobile networks (Ozturka et al., 2019; Nadembega et al., 2014), except very little works, like the work carried out in (Almeida, Azkune, 2018) in which the mobility prediction (for elementary actions in a limited space) has been applied to detect behavioral abnormalities for elderly people. Therefore, the challenge here consists in proposing solutions (models, algorithms...) for mobility prediction which

are applicable to other fields such as personalized recommendations systems.

On the other hand, although the prediction models proposed are oriented towards mobile networks, some of these models do not support certain types of networks such as UDNs which currently represent a new trend, such as 5G networks. UDNs are characterized by their high number of cells (base stations) (Zhang, Dai, 2018) and their management complexity (Samarakoon et al., 2016). This increases the complexity of certain models and prediction algorithms considerably (Zhang, Dai, 2018). Consequently, these models are not highly recommended to be applied to this type of networks because of the costs involved. Indeed, in the case of Markov chains-based models, the complexity of the transition matrix (main parameter in Markov models (Amirrudin et al., 2013a)) increases with the growth of the number of cells, in particular in case of hidden Markov model. Also, in the case of neural networks-based models (characterized by their computational complexity (Zhang, Dai, 2018)), complexity increases in case of a large number of cells (Zhang, Dai, 2018). Thus, proposing a solution that supports this type of network with lower complexity (cost) constitutes another challenge.

5.5 Challenge 5 - Evaluation of models

Evaluation of mobility prediction solutions can be done using simulations (Ozturka et al., 2019; Wang et al., 2019) often based on datasets containing data about individuals' movements. However, on one hand, these datasets may already be existing and the data they contain may not be adequate for the proposed solution, as for context-based solutions. On the other hand, some datasets set up for evaluations (Samaan, Karmouch, 2005; Göndör et al., 2013) are not very consistent in terms of the amount of data they contain (Göndör et al., 2013). In addition, datasets must often be divided into two parts: a first part for learning and a second part for validation which is based on the comparison between the second part of the dataset and the prediction results. For this, the datasets must be consistent enough so that they can be divided; otherwise, the evaluation may not be 100% correct (Göndör et al., 2013). A major challenge consists in the use or creation of a dataset which is the most complete regarding to the data necessary for the proper functioning of the proposed solution.

Another way of evaluation consists of a real evaluation (Chon et al., 2013), based on the participation of people who are available for real-time interaction (feedback), allowing to compare the results of mobility prediction with the real future individuals' movements. This approach allows for a more reliable evaluation (Göndör et al., 2013). To realize this kind of evaluation, the challenge consists in ensuring the participation of a large number of people. One way to achieve this objective can be the realization of a mobile application (Göndör et al., 2013) which is used by a group of people, and which allows the creation of a dataset by recording data necessary for prediction. These same people must also participate in the evaluation of the prediction results of their movements.

5.6 Challenge 6 - Confidentiality and privacy of users data

Mobility prediction solutions are mainly based on the historical data of individuals' movements, on their profile data, on their social relationships or even on their schedules. These data are directly linked to the privacy of individuals and need a certain degree of confidentiality. Indeed, the disclosure of personal data

related to an individual's movements or their schedules can have a negative influence on this individual. In other words, anyone who has information about an individual at disposal can harm the daily life of this individual by going from a simple tracking action to physical damage (theft, assault, etc.). Thus, confidentiality in the field of mobility prediction mainly concerns the confidentiality of the historical data of individuals as well as the information which can be deduced from this data.

In (Pan et al., 2013), the authors emphasized the confidentiality aspect of the trace data of individuals moving in a smart city and concluded that the disclosure of personal identity could occur during the collection, publication and use of trace data. With regard to collection, localization techniques may record user or device ID and cause risks. Location by GPS is more secure than GSM, WiFi, Bluetooth and RFID because centric servers do not need to know device IDs. Regarding the publication, personal identity could be inferred from published locations, in spite of having been removed/anonymized from the trace data records. In terms of usage, traces may expose unwanted privacy information to personalized services and applications. In such case, an anonymizing proxy is to be trusted to store, manage, and protect user locations, and to communicate between applications and users. The challenges are to keep fidelity of data for applications meanwhile protecting privacy (Pan et al., 2013).

6. CONCLUSION

In this paper, we've presented some key points on mobility prediction and related challenges, in order to give a guideline for the readers for new contributions. The first part of the paper was devoted to key concepts of mobility prediction through an overview about the mobility concepts and classifications, the data required for mobility prediction and some works dealing with mobility prediction. The second part focuses on a set of challenges, that are an open issues related to mobility prediction. These challenges were extracted from our deep literature study on the mobility prediction works. The second part can be very useful for a large community of researchers and developers because, to the best of our knowledge, there is no work that groups all these challenges at the same time.

From this work, we have noted that, as mobility data, contextual data are very important to predict mobility. Also, some issues such as predicting random mobility, based-context prediction, a long-term prediction and definition of appropriate datasets for evaluation of models are always open to new contributions. Mobility prediction is applied only to MNs domain. Consequently, it is interesting to apply it to other domains (e.g. recommendation systems). Also, no work has shown its interest for the confidentiality of individual mobility data. So, it would be interesting to explore this crucial point. In the future, we plan to propose a solution (model, method or algorithm) that tackles the main challenges mentioned.

REFERENCES

- Abowd, G. D., Dey, A. K., Brown, P. J. *et al.*, 1999. Towards a better understanding of context and context-awareness. In: *Int. symp. on handheld and ubiquitous computing*, Springer, Berlin, Heidelberg, 304-307.
- Abu-Ghazaleh, H., Alfa, A. S., 2009. Application of mobility prediction in wireless networks using markov renewal theory. *IEEE Transactions on Vehicular Technology*, 59, (2), 788-802.

- Almeida, A. Azkune, G., 2018. Predicting Human Behaviour with Recurrent Neural Networks. *Appl. Sci.*, 8 (2), 305.
- Amirrudin, N. A., Arif_n, S. H. S., A. Malik, N. N. N., Ghazali, N. E., 2013a. Mobility prediction via Markov model in LTE femtocell. *Int. J. Comput. Appl.*, 65(18), 40_44.
- Amirrudin, N. A., Arif_n, S. H. S., A. Malik, N. N. N., Ghazali, N. E., Dec. 2013b. User's mobility history-based mobility prediction in LTE femtocells network. In *Proc. IEEE Int. RF Microw. Conf. (RFM)*, 105_110.
- Anagnostopoulos, T., Anagnostopoulos, C., Hadjiefthymiades, S., 2012. Efficient location prediction in mobile cellular networks. *Int. Journal of Wireless Information Networks*, 19 (2), 97-111.
- Anisetti, M., Ardagna, C. A., Bellandi, V. *et al.*, 2011. Map-based location and tracking in multipath outdoor mobile networks. *IEEE Transactions on Wireless Communications*, 10 (3), 814-824.
- Benouaret, I., 2017. *Un système de recommandation contextuel et composite pour la visite personnalisée de sites culturels*. Thèse de doctorat.
- Boc, M., De Amorim, M.D., Fladenmuller, A., 2011. Near-zero triangular location through time-slotted mobility prediction. *Wireless Networks*, 17 (2), 465-478.
- Chamek, L., Daoui, M., Lalam, M., Nov. 2012. Prediction de mobilité basée sur la classification selon le profil. *Courrier du Savoir – N°14*, 113-117.
- Chon, Y., Talipov, E., Shin, H. *et al.*, 2013. SmartDC: Mobility prediction-based adaptive duty cycling for everyday location monitoring. *IEEE Transactions on Mobile Computing*, 13 (3), 512-525.
- Dash, M., Koo, K. K., Gomes, J. B., Krishnaswamy, S. P., Mar. 2015. Next place prediction by understanding mobility patterns. In *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops*, 469_474.
- Du, Y., Fan, J., Chen, J., 2011. Experimental analysis of user mobility pattern in mobile social networks. In: 2011 IEEE Wireless Communications and Networking Conference. IEEE, 1086-1090.
- Gambis, S., Killijian, M.-O., del Prado Cortez, M. N., 2012. Next place prediction using mobility Markov chains. In *Proc. 1st Workshop Meas., Privacy, Mobility*, p 3.
- Göndör, S., Uzun, A., Rohrmann, T. *et al.*, 2013. Predicting user mobility in mobile radio networks to proactively anticipate traffic hotspots. In: *2013 Int. Conf. on MOBILE Wireless MiddleWARE, Operating Systems, and Applications*. IEEE, 29-38.
- Jiang, J., Pan, C., Liu, H., Yang, G., Nov. 2016. Predicting Human Mobility Based on Location Data Modeled by Markov Chains. IEEE UPINLBS'2016, Shanghai, P.R. China.
- Liou, S.-C. Huang, Y.-M., 2005. Trajectory predictions in mobile networks. *Int. J. Inf. Technol.*, 11(11), 109_122.
- Liu, G., Maguire, G., 1996. A class of mobile motion prediction algorithms for wireless mobile computing and communications. *Mobile Networks and Applications*, 1 (2), 113-121.
- Liu, T., Bahl, P., Chlamtac, I., 1998. Mobility modeling, location tracking, and trajectory prediction in wireless ATM networks. *IEEE Journal on selected areas in communications*, 16 (6), 922-936.
- Lv, Q., Mei, Z., Qiao, Y., Zhong, Y., Lei, Z., Sep. 2014. Hidden Markov model based user mobility analysis in LTE network. In *Proc. Int. Symp. Wireless Pers. Multimedia Commun. (WPMC)*, 379_384.
- Lytrivis, P., Thomaidis, G., Tsogas, M. *et al.*, 2011. An advanced cooperative path prediction algorithm for safety applications in vehicular networks. *IEEE Trans. Intell. Transp. Syst.*, 12 (3), 669-679.
- Mcinerney, J., Zheng, J., Rogers, A., Jennings, N. R., 2013. Modelling heterogeneous location habits in human populations for location prediction under data sparsity. In *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 469-478.
- Nadembega, A., Hafid, A., Taleb, T., 2014. A destination and mobility path prediction scheme for mobile networks. *IEEE transactions on vehicular technology*, 64 (6), 2577-2590.
- Ozturka, M., Gogateb, M., Oniretia, O., Adeelb, A., Hussainb, A., Imrana, M. A., 2019. A novel deep learning driven, low-cost mobility prediction approach for 5G cellular networks: The case of the Control/Data Separation Architecture (CDSA). *Article in Neurocomputing*, 358, 479-489. (Preprint submitted to Elsevier Jan., 2019).
- Pan, G., Qi, G., Zhang, W. *et al.*, 2013. Trace analysis and mining for smart cities: issues, methods, and applications. *IEEE Communications Magazine*, 51 (6), 120-126.
- Parija, S., Ranjan, R. K., Sahu, P. K., Jan. 2013. Location prediction of mobility management using neural network techniques in cellular network. In *Proc. Int. Conf. Emerg. Trends VLSI, Embedded Syst., Nano Electron. Telecommun. Syst. (ICEVENT)*, 1_4.
- Qiao, S., Shen, D., Wang, X., Han, N., Zhu, W., Feb. 2015. A self-adaptive parameter selection trajectory prediction approach via hidden Markov models. *IEEE Trans. Intell. Transp. Syst.*, 16, (1), 284_296.
- Samaan, N. Karmouch, A., 2005. A mobility prediction architecture based on contextual knowledge and spatial conceptual maps. *IEEE transactions on mobile computing*, 4 (6), 537-551.
- Samarakoon, S., Bennis, M., Saad, W. *et al.*, 2016. Ultra dense small cell networks: Turning density into energy efficiency. *IEEE Journal on Selected Areas in Communications*, 34 (5), 1267-1280.
- Samuel Pierre, 2003. *Réseaux et systèmes informatiques mobiles : Fondements, architectures et applications*. Presses inter Polytechnique, 623 pages.
- Soh, W. S. Kim, H. S., Oct. 2006. A predictive bandwidth reservation scheme using mobile positioning and road topology information. *IEEE/ACM Trans. Netw.*, 14 (5), 1078_1091.
- Wang, H., Yang, Z., Shi, S., 2019. Next location prediction based on an adaboost-Markov model of mobile users. *Sensors*, 19 (6), p 1475.
- Wickramasuriya, D. S., Perumalla, C. A., Davaslioglu, K., Gitlin, R. D., Apr. 2017. Base station prediction and proactive mobility management in virtual cells using recurrent neural networks. In *Proc. IEEE Wireless Microw. Technol. Conf. (WAMICON)*, 1_6.
- Zhang, H., Dai, L., 2018. Mobility prediction: A survey on state-of-the-art schemes and future applications. *IEEE Access*, 7, 802-822.
- Zhang, Y., Hu, J., Dong, J. *et al.*, 2009. Location prediction model based on Bayesian network theory. In: *GLOBECOM 2009-2009 IEEE Global Telecommunications Conference*. IEEE, 1-6.
- Zheng, Y., Xie, X., Ma, W.-Y., *et al.*, 2010. GeoLife: A collaborative social networking service among user, location and trajectory. *IEEE Data Eng. Bull.*, 33 (2), 32-39
- Zimmermann, A., Lorenz, A. and Oppermann, R., 2007. An operational definition of context. In: *International and Interdisciplinary Conference on Modeling and Using Context*. Springer, Berlin, Heidelberg, 558-571.