



MOBILITY PREDICTION USING MOVEMENT, SEMANTIC AND SOCIAL DATA: A SURVEY

S.Sridhar raj¹, M.Nandhini²

Abstract—Mobility prediction of human and objects have a variety of application benefits such as tracking of a mobile device, vehicle route optimizing, human activity time optimization, etc., In order to perform these mobility predictions, large amount of collection and analysis of positioning data from sensors is required. These patterns are mined and predicted based on the movement, semantic and social data. In this paper, we survey about the various applications of mobility pattern prediction based on the type of data used and data mining techniques applied by categorizing the movement prediction of human and objects to next single or sequence of locations.

Keywords—Data mining, Mobility prediction, Spatial-Temporal data, Trajectory analysis, Pattern mining

I. INTRODUCTION

Data mining is one of the steps for future prediction and knowledge discovery. Due to the rapid development of technology with respect to software, hardware and automation systems, huge amount of data has been generated and stored in data marts and data warehouses. The traditional informal statistical methods and other such tools used to mine data are not sufficient to analyse this huge collection of data. Hence, an intelligent data analysis technique known as Knowledge discovery has been identified, so that knowledge or patterns can be extracted from various data sets.

The mobility prediction is defined as tracking the movement patterns of the human or object and predicting their future locations based on the mined patterns. It can be categorized into two types:

Trip-based, where aggregated mobility is considered (trips between different areas are predicted from the recorded data, and source destination matrices are generated);

Activity-based, where individual mobility is considered (each individual entity is locations to access, and based on their actions, trips are generated) [1].

Mobility prediction has attracted good research interest due to its vast applications like prediction of single or sequence of locations, mobility pattern behaviour for group or individual, urban service based applications, etc., It could be applied to any field such as government sectors, healthcare units, scientific domains and other similar domains, so that they are able to discover the patterns and predict results well in advance to safeguard the collapses which might occur in future. The best way to collect the movement data is through the GPS enabled devices especially mobile phones [2].

In Wireless Sensor Networks (WSN), the object mobility prediction has a variety of benefits like smart homes, enterprise applications, environmental monitoring, education and military hospital services [3]. In this work, we present a survey of movement prediction of human and objects based on next single or sequence of locations.

The remainder of the paper is organized as follows. Section II discusses the background, motivation and basic definitions. Section III presents the study on various applications on mobility predictions of both human and object. The comparative assessment of the study is presented in Section IV and Section V concludes the paper with future works.

II. BACKGROUND AND MOTIVATION

Predicting the future move of human or object helps us to reserve resources. Some of such movement prediction applications are Vehicle routing problem, social networking applications, reminder, aged people movement optimization, recommender systems and assistive devices. These applications are beneficial to the common people, commercial organizations, government agencies, health care units and safety critical systems.

The definitions of the movement data such as trajectory, temporal, spatial, semantic and social data are given below:

- A trajectory is the path made by the moving entity through the space where it moves [4].
- Temporal data is data that varies over time. A temporal data denotes the evolution of an object characteristic over a time interval [5].
- The spatial attributes of an entity consists of information representing spatial locations, e.g., longitude, latitude, shape and elevations [4].

¹ Department of Computer Science, Pondicherry University, Puducherry, India.

² Department of Computer Science, Pondicherry University, Puducherry, India.

- Semantic data denotes the contextual meaning and the social data involves various factors which can influence the movement [7].

Temporal data mining deals with the ordered events which has more than one dimension of time. Temporal mining can be differentiated into two categories; one is random discovery of relationships between temporal based events and the other is similar pattern mining based on time intervals. The time series is defined as series of timestamps in the data. Time series analysis is the analysis of discrete-time data in order to extract useful statistical data. The time series analysis also helps us in pattern matching, noise reduction and curve approximation with mathematical methods [6].

In temporal data mining, the mining process is identifying the relationships among timely events and also related casually. But, in non-temporal data mining, the mining process is discovery based on rules which forms the term “temporal rule discovery”. The temporal rule discovery respects the order of events which forms a sequence on the time axis. Temporal data mining can be performed on a sequence which gives rise to “sequence pattern mining”.

The above defined term, sequence pattern mining can also be interpreted as pattern discovery, since it mines and discovers new patterns. The other perspective of mining process is the group behaviour, which are clustered based on statistical properties.

Spatial data mining can also be stated as the multi dimensional form of temporal data mining. By increasing the number of dimensions, the numbers of techniques which can be applied are restricted. Thus, till date the spatial data mining considers the static techniques like clustering, association rules and characterisation [8].

Semantic data mining is the mining of data set by understanding the meaning and context of the data. Knowing the spatial and temporal context gives us trajectory information, whereas the semantic method gives the real meaning of the data [7].

Social data mining cannot be restricted with the finite set of parameters. Every factor which affects the movement either through internal or external factors have to be considered for mining. For example, the influence of mobile phones, neighbour interaction, etc., leads to distortion in the prediction accuracy. To address this inefficiency, user specific social data has to be mined [7].

Mobility prediction eventually assures us to optimize time and space by mining the historical data. These are various insights required to understand the mobility prediction applications which is discussed in the next section.

III. MOBILITY PREDICTION APPLICATIONS

The applications are grouped based on the location prediction task which gets broadly classified as human and object which is further classified as next or current location (predicts only a single future location) and sequence of locations. The goal, methodology, limitations and results of mobility prediction applications are discussed in this section:

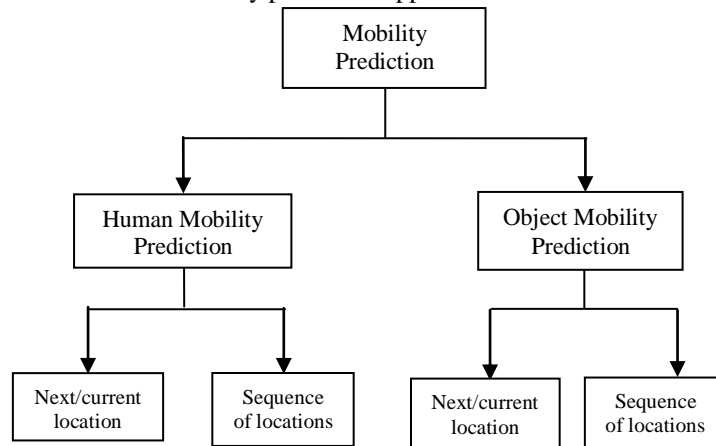


Figure 1. Categorization of mobility prediction applications

Predicting next or current location and sequence of locations is based on choosing the best one which is efficient in terms of cost, time and space.

A. Human mobility prediction

The movement of human differs in a large scale from objects. The factors considered to mine the human movement are huge when compared to objects. The movement of human can be influenced by both external and internal factors. Some of the applications which are developed to mine human movement patterns are discussed in this section:

1) Next or current location:

Mobile service systems are offering useful movement information through mobile devices. Based on dynamic user behaviour, mobile service systems have the capacity of effectively mining an intended request from huge data sets. The nearest location associated with the services requested by the user can be predicted and recommended by mining the historical movement behaviour patterns. For this purpose, a novel mining scheme [9] was proposed to mine the movement patterns using associated trees. It also presents an algorithm which counts the weighted support values from the matching candidate frequent patterns which improve efficiency and scalability.

In [10], the destination is understood and predicted through mined mobility patterns based on frequently visited location data. Activity transitions are more accurate than location transitions and hence Hidden Markov Model (HMM) is employed to estimate it. They use location semantics and movement patterns of a user to evaluate the probability of each activity in a supervised way. They provide a hierarchical model which predicts the next movement more accurately than other type of models.

In [11], they have presented a system based on user specific activities which analyses the movement prediction of the individual. The main aim of the prediction is to analyse the movements and categorize the locations with respect to user specific activity patterns. The prediction technique is performed by multi-agent systems. To estimate the future location of the user device, neural network and probability based prediction system are considered. Location prediction has been performed on analysing only 10percent of the movement.

The method presented in [9], mined the movement patterns using the traditional associated trees technique to suggest the nearest service location requested by the user, which was improved in [10] by considering activity transition and hierarchical model. The model defined in [11], mined user specific movement data with the help of multi-agents which further improves the prediction performance. All these models predict only the current and next location of human.

2) *Sequence of locations*

The model [12] predicts people's routines from the data extracted from mobile phones. It is implemented by a novel probabilistic approach based on location, variation in people's routines with respect to time and proximity information. The prediction is based on unsupervised discovery of the location and proximity acquired from the cell phone Bluetooth information.

In [13], they proposed a Markov model which is activity based on time. It is built to define and predict the human movement patterns. This model is a simple extension of mobility prediction algorithm which keeps track of activity transitions. A small university campus is considered to obtain the movement patterns and experiments are conducted based on these patterns. The model efficiently predicts the students' activities in the indoor university campus based on the jump methodology.

In [14], they describe new algorithms to mine and predict people's movement in an indoor environment. Movement patterns are mined from historical trajectory data, and these patterns are used to construct a probability tree which visually represents the frequent movements. They have conducted an experiment in a staff tearoom to capture the trajectory data of the staff and mine their movement patterns to construct a probability tree. The performance of prediction is determined based on the trajectory estimation strategies. Finally, the model predicts the sequence of movement of a staff which can be used for path optimization.

The above models are discussed in three different scenarios, where the first one [12] happens to be a general proximity problem subjected to any location, second [13] in a small university campus and the third [14] in a staff tea room. All the models predict the next sequence of movements of a person.

B. *Object movement pattern mining*

Unlike human movement pattern mining, objects have a easier way to mine with less factors influencing them. Some of the applications on object pattern mining are discussed here:

1) *Next or current location*

In [15], they have proposed a Heterogeneous Tracking Model (HTM), which mines the object moving patterns and also tracks it. In order to solve the dependency problems among object movements, they have used a variable memory Markov model. HTM is basically hierarchical in nature, which provides multi-resolution object moving patterns. This model predicts the movements of objects and uses the energy efficiently while tracking the object.

In [16] they have proposed an object tracking method in sensor networks through data mining. The traditional A priori algorithm is used to mine the association rules and then applied it to sensor networks applications. This data mining algorithm is performed over the historical movement data of the object to extract the association rules, which is used to predict the future location.

The model presented in [16] predicts the next location of the object accurately which increases the network lifetime and outperforms [15] in terms of energy efficiency and tracking accuracy.

2) *Sequence of locations*

The existing object mobility prediction methodologies focus mainly on geometric properties of trajectories. The geographic information and semantics are not considered for prediction. In [17], they have proposed a reverse engineering framework for

to mine and model the patterns of semantic trajectory data. A special type of relationship is required to model the conceptual schema of geographic databases due to trajectory data dependency problem. Therefore, they apply data mining to extract general trajectory patterns and these patterns are modelled for prediction based on the created geographic database schema.

In [18], the movement prediction is considered as classification problem by comparing the movement pattern of a certain object with the stored information. They have used Optimal Stopping Theory (OST) to avoid potential noise in movement pattern and to relax the slightly varying tasks for accurate sequence prediction.

The model presented in [19] concerns about the frequent movement of objects by combining the concepts of clustering and sequential pattern mining. The model considers both spatial and temporal data for the estimation of frequent patterns and cluster formation. The obtained clusters are linked in order to form the frequent trajectories. The efficiency of this method can be attained only when the cluster size is 0.3 units.

A sequential pattern fails when a new user’s movement has to be predicted and because of the similarity among the users, degradation is formed. Clustering and sequential pattern mining together helps in extracting the mobility rules from similar type of users in [20]. Clustering based sequential and sequential based clustering are two combinations of algorithms used to rectify the inadequacy of information and reduces random movement noises.

In [21], an efficient algorithm is proposed to mine the uncertain frequent patterns using novel data mining techniques. This model avoids the false positives and verifies the correctness of the mined patterns. The proposed method is a list based data structures and pruning techniques which allows efficient mining of a complete uncertain frequent pattern set. The advantage of this model is that, there are no pattern losses during the mining process and it outperforms the previously stated works [17, 18, 19, and 20].

The semantic trajectory model [17] mines the data with respect to semantic context; the OST in [18] adds accuracy to the prior works by removing the noise in the movement patterns. [19], [20] follows the combination of clustering and sequencing pattern mining to find the frequent trajectories and reduce random movement noises respectively. Outperforming all the previous works, the model [21] guarantees the correctness of the mined patterns and assures a better performance in terms of memory usage, runtime and scalability.

IV.COMPARISON OF ALL TECHNIQUES

In all the mining pattern techniques discussed above, they have used certain types of data to mine the patterns. The major types of data which influence the human movement are spatial, temporal and social data. Semantic data creates only less impact on the context of movement. The types of data used by the above applications are discussed in the table 1.

Table -1 Types of data used for movement pattern mining

Study	Spatial data	Temporal data	Semantic data	Social data
Human patterns – Next or current location prediction				
[9]	✓	✓	✗	✗
[10]	✓	✗	✓	✗
[11]	✓	✓	✗	✗
Sequence of locations				
[12]	✓	✓	✗	✗
[13]	✓	✗	✗	✗
[14]	✓	✗	✗	✗
Object patterns – Next or current location prediction				
[15]	✓	✓	✗	✗
[16]	✓	✗	✗	✗
Sequence of locations				
[17]	✓	✓	✓	✗
[18]	✓	✗	✗	✗
[19]	✓	✓	✗	✗
[20]	✓	✗	✗	✗
[21]	✓	✗	✗	✗

By observing the Table1, all the applications use spatial data. Obviously, it is the default data to be used when it comes to trajectory mining. Usage of temporal data increases the efficiency, but also brings in complexity. Only some works have considered the temporal data. The social data influences the people through their personal habits and social connectivity. As the technology grows, social data continues to create a greater impact in peoples’ routines.

Table -2 Various data mining techniques applied to the existing works

Study	Data pre-processing	Classification	Outlier detection	Cluster analysis	Prediction task	
					Next/Current location	Sequence of locations
Human Mobility Pattern prediction applications						
[9]	✗	✓	✗	✗	✓	✗
[10]	✓	✓	✗	✗	✓	✗
[11]	✗	✓	✗	✗	✓	✗
[12]	✗	✓	✗	✗	✓	✓
[13]	✗	✓	✗	✗	✓	✓
[14]	✗	✓	✗	✗	✓	✓
Object Mobility pattern prediction applications						
[15]	✗	✓	✗	✗	✓	✗
[16]	✗	✓	✗	✗	✓	✗
[17]	✗	✓	✗	✗	✓	✓
[18]	✓	✓	✗	✓	✓	✓
[19]	✗	✓	✗	✓	✓	✓
[20]	✗	✓	✗	✓	✓	✓
[21]	✗	✓	✓	✗	✓	✓

The Table 2 depicts the data mining techniques applied over the existing works. Each technique adds a level of strength for the mining process. Observe the data pre-processing column where most of the works fail to perform this technique which is necessary for accurate mining. It helps in reducing the redundancy and unwanted noises in the data before being mined. Secondly, the classification process is carried out in all the works which denotes how essential it is for the mining process. The core task of mining is classification where the system is trained to identify the safe and unsafe state. Then, the outlier detection phase is the detection of abnormal patterns. It is used in safety applications, where the pattern matching has to be done to detect the odd one out. The clustering mechanism is used to group the mined patterns into clusters which benefits in further processing. Finally, the prediction is carried out with all the previously stated processes. With respect to the movement prediction, there are two categories; one is prediction of next or current location. Given an object or human, it can give its current location and next location. But, second category which is the sequence pattern mining gives us the entire trip of the person or object. It is trivial that an application predicts the next location if it can predict the sequence of locations. Very few works have concentrated on the sequence pattern mining.

Considering the limitations stated in the above comparative assessment, it is obvious that a new mobility prediction model comprising the spatial, temporal, semantic and social data can yield better accuracy. Data pre-processing is an important level of data mining process which is rarely addressed in majority of the above discussed works. It is observed that, considering the pre-processing phase to develop the prediction model improves the efficiency of the latter techniques such as classification, clustering, outlier detection and prediction mechanisms. Performing anomaly detection of abnormal patterns and clustering of the patterns adds efficiency to the proposed model.

The performance of the new idea which is derived from the limitations can be evaluated with the human movement prediction applications such as military services, aged care, etc., and object movement prediction applications such as vehicle routing problem, sensor node tracking and social networking applications.

V. CONCLUSION AND FUTURE DIRECTIONS

Various techniques used for movement pattern mining are studied in this paper. The movement prediction of both human and object are considered for the survey. They are compared with each other based on the type of data used and the data mining techniques applied over it to acquire the final prediction results. The study clearly shows that very less number of researches has been carried out in predicting the entire routine of human which is a tedious task. In future, exploration of social influence on movement, considering the deep learning algorithms to train the classifier can be carried over.

REFERENCES

- [1] F. Calabrese, G. D. Lorenzo and C. Ratti, "Human mobility prediction based on individual and collective geographical preferences," in Proceedings of the 13th International IEEE Annual Conference on Intelligent Transportation Systems, pp.312-317, Sept 19-22. 2010.
- [2] Q. Ye, L. Chen and G. Chen, "Predict Personal Continuous Route," in Proceedings of the 11th International IEEE Conference on Intelligent Transportation Systems, pp.587-592, Oct 12-15. 2008.
- [3] T.S. Chen, C.H.Wu and J. Jee, "Object tracking by mining movement trajectories in wireless sensor networks," IEEE/IFIP NOMS 2014 - IEEE/IFIP Network Operations and Management Symposium: Management in a Software Defined World, 2014.
- [4] S. Shekhar, R. R. Vatsavai and M. Celik, "Spatial and spatiotemporal data mining: Recent advances," Next generation data mining, 2008.
- [5] S. laxman and P. S. Sastry, "A survey of temporal data mining," Sadhana, vol.31, pp.173-198, April 2006.

- [6] J. F. Roddick and B. G. Lees, "Paradigms for Spatial and Spatio-temporal Data Mining," Geographic Data Mining and Knowledge Discovery. Taylor and Francis. Research Monographs in Geographic Information Systems, pp. 1-14, 2001.
- [7] J. J. C. Ying, W.C. Lee, T.C. Weng and V. S. Tseng, "Semantic Trajectory Mining for location prediction," Proceedings of the 19th ACM SIGSPATIAL International Geographic Information Systems, pp.34-43, Nov 01 -04. 2011.
- [8] J. F. Roddick and M. Spiliopoulou, "A Bibliography of Temporal, Spatial and Spatio-Temporal Data Mining Research," SIGKDD Explorations, vol. 1(1), pp.34-38, June. 1999.
- [9] T. S. Chen, Y.S. Chou and T. C. Chen, "Mining user movement behavior patterns in a mobile service environment," IEEE Transactions on Systems, Man and Cybernetics: Systems, vol.42(1), pp.87-101,Jan.2012.
- [10] W. Huang, M. Li, W. Hu, G. Song and K. Xie, "Hierarchical destination prediction based on GPS history," Proceedings - 2013 10th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2013, pp.972-977, 2013.
- [11] M. Vukovic and D. Jevtic, "Agent-based movement analysis and location prediction in cellular networks," Procedia Computer Science, vol. 60(1), pp.517-526, 2015.
- [12] K. Farrahi and D. G. Perez, "Discovering human routines from cell phone data with topic models," Proceedings - International Symposium on Wearable Computers, ISWC, pp.29-32, 2008.
- [13] J.Kolodziej, L. Wang, N. Ghani, S. U. Khan,N. Min-Allah, S. A. Madani, and H. Li, "An application of Markov jump process model for activity-based indoor Mobility Prediction in wireless networks," Proceedings - 9th International Conference on Frontiers of Information Technology, FIT, pp.51-56, 2011.
- [14] L. D. M. Lam, A. Tang, and J. Grundy, "Predicting indoor spatial movement using data mining and movement patterns," IEEE International Conference on Big Data and Smart Computing, pp. 223-230, 2017.
- [15] W. C. Peng, Y.Z. Ko, and W.C. Lee, "On mining moving patterns for object tracking sensor networks," Proceedings - IEEE International Conference on Mobile Data Management, pp. 0-3, 2006.
- [16] W. H. Liao, K. C. Chang, S. P. Kedia, "An object tracking scheme for wireless sensor networks using data mining mechanism," IEEE Network Operations and Management Symposium, pp.526-529, 2012.
- [17] L. O. Alvares, V. Bogorny, J. A. F. de Macedo, B. Moelans and S. Spaccapietra, "Dynamic modeling of trajectory patterns using data mining and reverse engineering," Tutorials, Posters, Panels and Industrial Contributions at the 26th International Conference on Conceptual Modeling, vol. 83, pp.149-154, 2007.
- [18] C. Anagnostopoulos and S.Hadjiefthymiades, "Intelligent Trajectory Classification for Improved Movement Prediction" IEEE Transactions On Systems, Man, And Cybernetics: Systems, vol. 44, No. 10, Oct. 2014
- [19] A.A.Shaw and N. P. Gopalan, "Finding frequent trajectories by clustering and sequential pattern mining," Journal of Traffic and Transportation Engineering (English Edition), vol. 1(6), pp.393-403, 2014.
- [20] T. V. T. Duong and D. Q. Tran, "A fusion of data mining techniques for predicting movement of mobile users," Journal of Communications and Networks, vol. 17(6), pp.568-581, Dec. 2015.
- [21] G. Lee and U. Yun, "A new efficient approach for mining uncertain frequent patterns using minimum data structure without false positives," Future Generation Computer Systems, vol. 68, pp.89-110, 2016.