A Survey of Location Prediction Using Trajectory Mining

B.A. Sabarish, R. Karthi and T. Gireeshkumar

Abstract This paper is a research and analysis on the prediction of location of moving objects that gained popularity over the years. Trajectory specifies the path of the movement of any object. There is an increase in the number of applications using the location-based services (LBS), which needs to know the location of moving objects where trajectory mining plays a vital role. Trajectory mining techniques use the geographical location, semantics, and properties of the moving object to predict the location and behavior of the object. This paper analyses the various strategies in the process of making prediction of future location and constructing the trajectory pattern. The analyses of various mechanisms are done based on various factors including accuracy and ability to predict the distant future. Location prediction problem can be with known reference points and unknown reference points, and semantic-based prediction gives an accurate result whereas the probability-based prediction for unknown reference points.

Keywords Location-based services · HMM · Personal communication system · GMPMINE and cluster ensemble algorithm · Trajectory mining algorithms

B.A. Sabarish (✉)
Department of Information Technology, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: sabarishpm@gmail.com

R. Karthi
Department of Computer Science Engineering, Asian College of Engineering and Technology, Coimbatore, India
e-mail: karthiamrita@gmail.com

T. Gireeshkumar
Department of Cyber Security, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: gireeshkumart@gmail.com

L.P. Suresh et al. (eds.), *Artificial Intelligence and Evolutionary Algorithms in Engineering Systems*, Advances in Intelligent Systems and Computing 324,
DOI 10.1007/978-81-322-2126-5_14

1 Introduction

Moving objects include vehicles, human, and mobile devices, which traverse across the geographical region (complete trajectory). Trajectory is the path of moving object traverse along and can be represented with the reference points or position in the path. It is essential to know the approximate location of the objects to be known to provide the location-based services (LBS). Many methods predict the location using the linear functions, but practically it follows a nonlinear dynamic function. Prediction can be classified as personal-based prediction and group-based prediction. Personal-based prediction involves the process of collecting information about the individuals independent of each other. This process produces unique trajectory pattern [1] for every individual, and this increases the number of trajectory pattern and complexity. In the other case, group-based mining process identifies the common trajectories and clusters the objects together and creates a general trajectory pattern. It provides a better result because the human tends to move in crowd than as individual. The issue arises in group-based mining is that there may be a chance of leaving out some interesting pattern when it is generalized into groups.

Figure 1 represents the general architecture of the trajectory mining process. The process starts from the collection of trajectory data and applying data mining techniques to gain the information and analyze the association between them. The gained knowledge can be used to predict the location behavior of the objects with the help of various prediction techniques including the statistical and probabilistic approaches (HMM), semantic-based prediction of various locations.

1.1 Personal-Based Prediction

In personal communication system (PCS), movement of mobile users is recorded in the database and service provider maintains it. The resources in communication can be dynamically allocated if the movement of the objects is predicted, which can improve the effective utilization of resources. Future prediction of mobile users can be easily done by efficient processing of their location-dependent queries regarding hotel and health care application. In the case of analyzing the individual movement pattern, mining can use the property of sequential data mining process. Individual movement will follow a pre-defined sequence of association rules, which can differ for working day or holiday or functions. For example, the movement of a human on

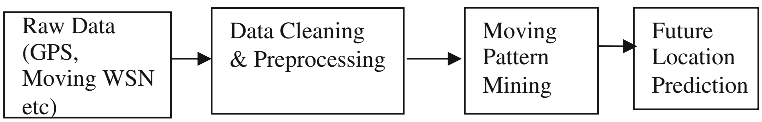

Fig. 1 General architecture of moving objects trajectory mining

a working day will be like Home → Bus Stop → Office → Bus stop → Home, etc. From this individual, mining information also the groups can be formed using the spatiotemporal attributes available in the historical data. The prediction accuracy again depends on the historical data collected and processed before the prediction process [2].

1.2 Group-Based Prediction

"Group-based data mining" is a data mining method involving the search for the association between individuals. Groups are formed based on the common behavior of the moving objects; it can be geographically closer or stay together for meaningful duration of time. Group mining is tedious when compared to individual pattern mining because the group behavior tends to have loss of information due to generalization. Normally, the process of finding frequency-pattern prediction involves GMPMINE and Cluster Ensemble algorithm for prediction of group movement patterns. In which, GMPMINE algorithm extracts the local association rules in the groups formed and CE algorithm puts the local association rules together to improve the association rule mining process. In Group-based prediction, the accuracy is calculated using punctual score and path score. Punctual score is the measure of relation of the moving object with respect to the region. Path score is an aggregate of the punctual score of all nodes along the path.

This paper is organized as follows. Section 2 describes motivation for this problem; Sect. 3 introduces terminology and needed preliminaries used in this paper. Section 4 gives some introduction about the trajectory mining techniques and analysis of various methodologies. Finally, Sect. 5 presents our conclusions.

2 Motivation

Predicting the future locations motivates the research issues in trajectory because of the dynamic behavior of the objects. Prediction of popular locations where most people visit leads to loss of information and leads to imbalanced data problem. And these prediction algorithms mainly predict only when the previous location (trajectory prefix) is known, which leads to loss of recall predictions. The performance of the prediction algorithm is affected by an increase in the number of moving objects, trade-off between the speed of prediction and accuracy, and ability to predict the distant future. The quality can be increased by means of partitioning and following client server architecture. The trajectory can be partitioned and clustered then prediction can be done at the cluster level. Method of client server architecture is proposed in which each individual object measures its own movement and server indexes the location at various level and uses queries using the filtering mechanism. Existing space partitioning approaches leads to two major issues answer-loss problem and granularity problem. The moving objects tend to move in groups, which leads to data-loss problem and representation semantics of various reference points. The success of the LBS is based on area location information in the particular time stamps accurately. It will be a challenge to predict the behavior of the system, if the object movement is fast and dynamic.

3 Preliminaries

3.1 Movement Pattern Mining

Spatial and temporal relations and similarities can be identified from the trajectory dataset which is normally represented as a sequence of reference points. The trajectories point can be compared against the prototypes, which are usually generated by the available information and previous history. Classification of usual and unusual behavior is done based on the similarity-level measure with the remaining behaviors. Peng et al. introduced a prediction model based on the incremental model of predicting and identifying the mobile user, which can be used for resource allocation problem in limited resource-based networks.

3.1.1 Periodic Pattern Mining

Periodic pattern means identifying the repetition or replica of the pattern happening after some duration of time. For example, the visit of tourists, migration of birds and animals during a particular season for the year and used to find the behavior of the moving objects. It can be used to identify the peak hours of traffic during weekdays and weekends. From this periodic identification the prediction of the movement, a prediction can be done on the expectation of crowd and population details, etc. The main issue is the selection of appropriate, optimum period for predicting the behavior, which may range from an hour to a year.

3.2 Trajectory Clustering

Clustering of mobile objects has gained importance because of its dynamic behavior of movement. Trajectory clustering is the process of grouping similar paths. While comparing two trajectories, the middle some reference points (sub-trajectory) can be similar but not the whole path. This sub-trajectory has been identified as the impact sub-trajectories since it is used by many trajectories. These sub-trajectories can be used for specific area analysis. Trajectory clustering algorithm (TRACLUS) is proposed by Lee et al. [3] which is partition and group framework.

TRACLUS is divided into two phases: partitioning and grouping. In partitioning, each trajectory is divided into regions using the line segments. In grouping, the line segments are represented using minimum description length. Then, similar line segments should be grouped together using DBSCAN algorithm. TRACLUS measures the accuracy in terms of the preciseness and conciseness properties. Trajectory can be classified based on the features of the regions identified [3, 4].

Since the trajectory mining process is dynamic, the size of data will be huge so it will be nice option to choose the sample and approximation/summarization techniques to solve the problem of scalability. Makris et al. proposed a machine learning-based approach to identify the usual behavior of the user. The concept of route is used, which in turn has identified by the pair of source and destination with some control points in the middle of the trajectory. Similarity can be measured by comparing the trajectory with the route. If a new trajectory is found, then it cannot be directly classified as usual behavior it will monitored whether that trajectory is used most often based on that a new route is created or an unusual route is identified [5].

4 Trajectory Mining Algorithms

Trajectory mining is the process of mining the path traversed by various objects and applies the data mining principles to identify the frequent matching pattern and predicts the path the nodes may travel [6].

4.1 Prediction Using Hidden Markov Models

Predicting the future locations with HMM proposed by Wesley Mathew et al., implemented the prediction model in the GeoLife project. Wesley proposed a hybrid model using the HMM, which clusters the available locations in the trajectory and different HMM models have been created for each cluster to train the system to predict the future location and measuring the similarity of the patterns. The previously analyzed trajectories are clustered according to their time of occurrence, and each is trained using different HMMs to predict the near future location. The geographical data are in the continuous location, which is converted into discrete specific for the regions which in turn used as the states. Each state will be associated with the probability distribution function for all the possible transitions. It helps in identifying the next location with higher probability. It provides an accuracy of 13 %. Osamma et al. proposed a template matching strategy, which compares the available patterns available in the prefix [7].

Hoyoung et al. proposed a novel method to predict the future location using the HMM. Hoyoung introduced a trajectory pattern model that describes an object's movement patterns based on hidden Markov process. It consists of a set of N frequent regions, each of which is associated with a set of M possible partitioned cells.

Cells are classified as observable states and hidden states. The discovered frequent regions are marked as hidden states and others as observable states. The probability of each observation sequence is calculated, which compares the current trend movement against historical data. It should able to update the current trend along with the historical data of the available movement patterns identified. Accuracy of the prediction depends on the level of granularity of the information. To extract the frequent regions, a periodical mining method is used. The period of the mining data depends on the type of application, for example, animal movement is for a year, and human movement prediction can be for a day or an hour. The complete trajectory has to be divided to the equal number of periods, which is calculated and clustered together. Then, trajectory prediction model is constructed with the help of HMM. It identifies available states (N), observation Symbols in each state (M), initial state (π), state transition probabilities (A), and observation symbol probabilities (B). Using the state transition and observation symbol, probabilities and the initial state predict the future using the Baun–Welch algorithm [8].

4.2 Prediction Using String Matching Algorithms

In the trajectory mining using LCS, each reference point is considered as a character in a string. The prediction is done using the prefix string, i.e., the trajectory that has come across to reach the present state. Path similarity is measured in terms of similarity and importance. It should be non-overlapping and identical paths.

Banerjee and Ghosh [9] proposed a variation of LCS that is applied in Web usage mining which in turn can be applied to trajectory mining with slight variation. It uses the weighted LCS, which assigns weight value to the various reference points available on the trajectory based on the visits made to the reference point with respect to the available trajectory. Similarity graph can be constructed with the help of min-cut and balancing algorithm. Then, the cluster can be formed using the concept clustering algorithm. Ghosh specifies the tree construction start with the process of identifying first-level nodes, which constitutes to be the frequently visited reference points.

4.3 Semantic-Based Prediction Methods

Semantic trajectory mining is developed to improve the efficiency and using the meaningful movement of a human in order to trace and predict. It starts with identification of stay points and calculating the support and confidence value of each stay point with respect to another. Stay points are identified using the time spend on the particular location by human before making a decision to move or divert, etc. Stay points may serve as decision-making points. The trajectories can be presented with as a sequence of stay points. Each stay point is named with a semantic name for representation and improves the mining process to gain the knowledge about the user prediction. It represents the path like <home, {bank, park}…>…, it represents the trajectory that follows the path of home, bank, and park. It specifies the information that bank and park go together which has the high support/confidence value. Support/confidence value is calculated using the conditional probability and data mining principles. Josh Jia-Ching Ying introduced a framework Seman Predict to evaluate the users' next location which may be either online or off-line. Off-line uses the notion of the stay location information to represent the user movement behavior. From the stay point information extracted the individual user's trajectory information can be identified (i.e) Semantic Trajectory similarity measure [10].

4.4 Pattern Matching Algorithms-Based Prediction Models

Monreale et al. [11] proposed a WhereNext predictor algorithm to predict the next location in the trajectory using the previously identified trajectory pattern (T-pattern). T-pattern is the common behavior of group users in space and time, which consists of node identifier, region identifier (spatial component), support value, and children for the node. They proposed a four step approach to predict the future location: (i) data selection, (ii) local models extraction, (iii) T-pattern tree building, and (iv) prediction. WhereNext represented the trajectory or spatiotemporal sequence is represented with the help of triples (x_i, y_i, t_i), which corresponds to the location (x, y) in the time t. Trajectory pattern is an efficient algorithm to find out the frequent visiting location sequences using the threshold values including minimum support (σ) and temporal tolerance value (τ). Accuracy of this algorithm is analyzed using posterior analysis to calculate the average error rate, spatial and data coverage.

Prediction is done for three different scenarios

(i) WhereNext_{r-1} intersects the region of the Node r.
(ii) WhereNext_{r-1} enlarged by temporal tolerance t_h intersects the region of the node r.
(iii) Does not intersect the region even after extending t_h.

Best matching path is identified by means of maximum path score and easy admissible prediction of future locations. WhereNext algorithm fails in the following occasions mainly when the length of the trajectory is lengthier than all other patterns and when it is distance from T-pattern (spatial and temporal distance).

Morzy [12] introduced a new approach by identifying the association rules using a modified apriori algorithm and uses PrefixSpan algorithm for predicting the future location. These models identify the frequent matching item set in the trajectories. Based on the frequent patterns, a classifier can be constructed. The construction creation involves three major steps mainly feature generation, feature selection, and modeling the classifier based on the features extracted in previous stages. This trajectories identified should be generalized but that cannot be done on a raw trajectory.

Morzy suggested dividing all the trajectories in equal-sized squares, which is named as cell and has four edges. Each cell is identified by the coordinates <*x*, *y*>. When an object moves from one cell to the other, it crosses the edges. The edges can be either vertical (left/right) or horizontal (north or south).

Trajectories will be represented as the sequence of the edges it comes across, and the length of the trajectory is the number of edges it has come through. The trajectory can be maximal when the path does not match with any other patterns. Trajectories are divided into head and tail concatenation of both head and tail will lead to trajectory information. Support value for each trajectory is calculated and frequent trajectories in the trajectory. Frequent trajectories will be converted to a movement rule. If the movement is from T1 $\rightarrow$ T2 where T1 and T2 are adjacent directories, then T1 $\oplus$ T2 forms the frequent trajectory. T2 is named as head of the rule and T1 as the tail of the rule [12].

Morzy [12] proposed a way to decompose the location prediction problem into two sub problems:

- Discover movement rules with support and confidence greater than user-defined thresholds of min-sup and min-conf,
- Match the movement rules against the trajectory of a moving object for which the current location is to be determined.

5 Conclusion

This paper analyzes the various aspects of the trajectory mining algorithms. The trajectory-based algorithms are grouped based on the techniques; it uses to mine the trajectory information. Mainly they are based on probabilistic HMM models and string matching algorithms. By comparing the various techniques and algorithm, semantic-based algorithm provide better results for the trajectory with known and fixed reference points. The probabilistic and string matching algorithm-based prediction models can be used for unknown reference points, which provide a low accuracy. By analyzing the HMM and string matching algorithms, the string matching algorithms provide slightly better performance than the HMM-based models because most of the HMM models are designed specific for the local cluster, which changes dynamically.

References

1. J. Yang, M. Hu, TrajPattern, Mining Sequential Patterns from Imprecise Trajectories of Mobile Objects, EDBT 2006, pp. 664–681
2. L. Chen, M. Lv, Q. Ye, G. Chen, J. Woodward, A personal route prediction system based on trajectory data mining. Inf. Sci. **181**, 1264–1284 (2011)

3. J.G. Lee, J. Han, X. Li, H. Gonzalez, J.G. Lee, J. Han, X. Li, H. Gonzalez, TraClass: trajectory classification using hierarchical region-based and trajectory-based clustering. J. Proc. VLDB Endowment **1**(1), 1081–1094 (2008)
4. J.G. Lee, J. Han, K.Y. Whang, Trajectory clustering: a partition-and-group framework, in *Proceedings of the 2007 ACM SIGMOD International Conference on Management of data*, pp. 593–604
5. O. Ossama, H.M.O. Mokhtar, M.E. El-Sharkawi, An extended k-means technique for clustering moving objects. Egypt. Inf. J. **12**, 45–51 (2011)
6. F. Giannotti, M. Nanni, D. Pedreschi, F. Pinelli, Trajectory Pattern Mining, KDD'07 (2007) pp. 330–339
7. W. Mathew, R. Raposo, B. Martins, Predicting Future Locations With Hidden Markov Models Ubicomp (2012) pp. 911–918
8. H. Jeung, H.T. Shen, X. Zhou, *Mining Trajectory Patterns Using Hidden Markov Models* (Springer, Berlin, 2007), pp. 470–480
9. A. Banerjee, J. Ghosh, Characterizing visitors to a website across multiple sessions, in *Proceedings of the National Science Foundation (NSF) Workshop on Next Generation Data Mining*, pp. 218–227
10. J.J.C. Ying, W.C. Lee, T.C. Weng, V.S. Tseng, Semantic Trajectory Mining for Location Prediction GIS (2011), pp. 34–43
11. A. Monreale, F. Pinelli, R. Trasarti, F. Giannotti, in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *KDD'09* (2009) pp. 637–646
12. M. Morzy, Mining Frequent Trajectories of Moving Objects for Location Prediction MLDM'07, pp. 667–680
13. Z. Li, M. Ji, J.G. Lee, L.A. Tang, Y. Yu, J. Han, R. Kays, MoveMine: Mining Moving Object Databases SIGMOD10
14. J.G. Lee, J. Han, K.Y. Whang, Trajectory clustering: a partition-and-group framework, in *SIGMOD'07*
15. J.G. Lee, J. Han, X. Li, Trajectory outlier detection: a partition-and-detect framework, in *ICDE'08*
16. F. Tung, J.S. Zelek, D.A. Clausi, Goal-based trajectory analysis for unusual behaviour detection in intelligent surveillance. Image Vis. Comput. **29**, 230–240 (2011)
17. Y. Wang, E.P. Lim, S.Y. Hwang, Efficient mining of group patterns from user movement data. Data Knowl. Eng. **57**, 240–282 (2006)
18. H.P. Tsai, D.N. Yang, M.S. Chen, Mining group movement patterns for tracking moving objects efficiently. IEEE Trans. Knowl. Data Eng. **23**(2), 266–281 (2011)
19. G. Yavas, D. Katsaros, O. Ulusoy, Y. Manolopoulos, A data mining approach for location prediction in mobile environments. Data Knowl. Eng. **54**, 121–146 (2005)
20. T.H. Cormen, C.E. Leiserson, R.L. Rivest, *Introduction to Algorithms* (MIT Press, Cambridge, 1990)
21. I.V. Cadez, S. Gaffney, P. Smyth, A general probabilistic framework for clustering individuals and objects, in *Proceedings of Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2000) pp. 140–149