www.ijecs.in
International Journal Of Engineering And Computer Science ISSN: 2319-7242
Volume 5 Issue 12 Dec. 2016, Page No. 19532-19635

A Survey on Destination Prediction Using Trajectory Data Mining Technique

Banupriya C S[1], Dr.V.Vijaya Chamundeeswari[2]

Final Year Student, M.E Computer Sciences and Engineering,
Velammal Engineering College, Chennai
priyasitas@gmail.com
Head of the Department, M.E Computer Sciences and Engineering,
Velammal Engineering College, Chennai
vijaychamu@gmail.com

Abstract - ***Destination prediction technique involves tracking the trace of the moving objects with the positioning services like GPS equipped devices. Destination prediction is also called path discovery to find most suitable path that satisfy predefined objective between source and destination. Trajectory is represented by a sequence of time stamped geographical location. Trajectories provide intelligence to estimate, compare and construct candidate routes by historical road network. All trajectories are decomposed into sub-trajectories and then synthesized-trajectories are generated by connecting sub-trajectories together. Moving objects can be human being, animals, vehicles and even natural phenomenon (e.g., Hurricanes). Mobility pattern of the user is predicted using next check-in data. Prediction features that exploit different information dimensions about users based on venue prediction.***

KEYWORDS: Trajectory database, Destination prediction, Mobility pattern, Trajectory compression.

I. INTRODUCTION

Data mining is the process which is discovering pattern in large set of data involves application of artificial intelligence, machine learning, statistics and database system. It is to extract the information from dataset and transform into structured format for further use. Trajectory data mining are a subset of spatio-temporal data with spatial dimensions. Trajectory is a continuous mapping between locations. Destination prediction technique involves in tracking the trace of the moving objects with the help of positioning services like GPS equipped device [9]. Moving objects can be Human being, Animals, Vehicles and even Natural Phenomenon (e.g.: Hurricanes). Destination prediction is also called path. A Trajectory is represented by a sequence of time stamped geographical location. Trajectories provide intelligence to estimate, compare and construct candidate routes by historical road network. All trajectories are decomposed into sub-trajectories and then synthesized trajectories are generated by connecting sub-trajectories together. Mobility pattern of the user is predicted using next check-in data [2] .Prediction features that exploit different information dimensions about users based on venue prediction.

II. LITERARY SURVEY

Aiden Nibali and Zhen He [1] proposed the Trajic process for compressing the vast amount of the trajectory data. The common approaches for compressing are line generalization and delta based approaches. In line generalization fix the trajectories in the series of line segments i.e. approximate trajectory. In delta compression store the data points initially and then store the remaining data points as the sequence of successive delta. Trajic has the major component *predictor* which takes the previous sequence of approximation points to generate the next predicted points, *Residual makers* is the storage of data between the original point and the predicted points where residuals are stored and reconstruct the trajectory. To get the accuracy of the trajectory data we use *approximate points* and finally the collected trajectory residual is calculated before encoding where the *leading zero algorithm* uses the statistics of entire sequence of residual to find the optimal encoding.

Anastasios Noulas, Salvatore Scellato, Neal Lathia, and Cecilia Mascolo [2] proposed the user mobility feature of next place prediction provides unprecedented opportunity to collect fine grained spatio-temporal data about the place user visit based on the Location Based Services. Based on the set of prediction features information is exploited between types of places, mobility flows between venue and spatio-temporal characteristics of check-in pattern. Next check in problem aim is to predict the exact place a user will visit next given historical data and current location such as historical visits or social ties. The ranking is made by assessing the predictability of individual user i.e. prediction. Venue prediction is based on the mobility prediction features, the *user mobility features* check-ins generated by the user under prediction or social networks. Another source of behaviour is based on historical behaviour based on number of check-ins, in this way different categories of places (coffee shop, cinema, temple etc…) are identified and ranked. *Global mobility feature* is based on geographical distance activity transition, place transition, rank distance. Temporal feature capture the user activity based on

DOI: 10.18535/ijecs/v5i12.67

categories of places user visits and also the temporal pattern of visits to specific places.

Costas Panagiotakis, Nikos Pelekis et al. [3] proposed the segmentation and sampling of moving object and browse their spatio temporal content. Trajectory segmentation is based on the representativeness of sub-trajectories in the Moving Object Database(MOD).Global voting algorithm is based on local density and trajectory similarity where voting is used to detect the representative paths and sub-trajectories. Segmentation algorithm estimates the number of partitions into sub-trajectories and identifies the homogeneous partition with the representativeness irregular to their space complexity. Segmentation sampling algorithm applied over the partition provides the most representativeness sub-trajectory of MOD.

Han Su, Kai Zheng et al. [4] proposed the calibrating trajectory for the spatio-temporal similarity analysis described by the movement history of moving object. Trajectories are the continuous function mapping between time and space. Trajectory Calibration process is necessary before raw trajectory data can be used for the subsequent analysis to transform the set of heterogeneous trajectory dataset with the unified sampling. Calibration process is of geometry based and model based. In geometry based uses the spatial relation between the trajectories and anchor points. Anchor point is a stable and independent of the trajectory data source refer to the geographical object such as Point of Interest (POI). Model based calibration exploits the correlation between anchor points from histories and the calibration approach.

Han Su, Kai Zeng Kai Zeng et al.[5] proposed the partition and summarization approach of trajectory data to understand the raw trajectory automatically by and described in short text. In partition phase, set of feature is defined for trajectory segments and finding the optimal partition and make the segments homogeneous as possible. In summarization phase, the historical behaviour of the same route is compared with the common behaviour and generated by short text. Raw trajectory is represented by the discrete sequence of location sampled from moving objects continuously. Feature extraction are classified based on *routing feature* (describe where the moving object travel) based on the grades of the road, road width, direction and *moving feature* (describe how the moving object travel) based on the speed, number of stay points and number of U-turn. Trajectories collected from the different source has different format and translated into short text for easy storing and communication.

Lyu-Chao, Xin-Hua Jiang, Fu-Min Zou et.al [6] proposed the latent semantic information mining for the trajectory data is to solve problem of structuring the spatio-temporal data. Trajectory data composed of latitude, longitude and time information. By retrieving the latent information fully, several information can be provided for road planning optimization. Initially Vector Space Model (VSM) is mainly for structuring the trajectory data i.e. converting irregular, unstructured trajectory data into structured data. Trajectory data in spatial grids which is represented by rows and column. To simply the trajectory representation and avoid difficulties the redundant grid is deleted from the continuous grid. By defining the network space the dimension of the space is greatly compressed and improves the performance efficiently. Singular Value Decomposition reflects the context information between the movement behaviour. Latent semantic information provides a feasible solution for mining the spatio-temporal trajectory data and used in several applications.

Nikos Pelekis, Ioannis Kopanakis, Costas Panagiotakis, and Yannis Theodoridis [7] proposed the unsupervised trajectory sampling is efficient sampling trajectory database for mobility data mining purpose. *Symbolic vector representation* based voting techniques is transformed into continuous function that describe representativeness is relaxed by merging algorithm maximal representative of each trajectory. Vanishes the jerky movement, speed up computation and turn lossless in mobility pattern. *Merging algorithm* identifies the maximal time period wherein the mobility pattern of each trajectory is augmented in terms of representativeness/ uniform behaviour of representativeness. *Sampling algorithm* selects subset of a trajectory database in unsupervised way to encapsulate the original trajectory data.

Yanhua Li et al. [8] proposed the sampling big trajectory data increase the scale of spatio-temporal trajectory data. Trajectory aggregate query aims to retrieve the statistics of trajectory passing a user specified spatio-temporal region. Random Index Sampling (RIS) algorithm estimates the guaranteed error bounds to sample the leaf node in spatio temporal query and design the unbiased estimator. Concurrent Random Index Sampling (CRIS) algorithm is to process a number of trajectory aggregate queries arriving concurrently with overlapping spatio-temporal query to improve the system scalability.

Zhenni Feng and Yanmin Zhu [9] proposed the application and its techniques in trajectory data mining, track the moving object based on GPS (Global Position System) -enabled devices which is represented by a sequence of time stamped geographical location. The various applications of trajectory data mining e.g. path discovery, location/destination prediction, movement behaviour analysis, and so on. Moving objects can be Human being, Animals, Vehicles and even Natural Phenomenon (e.g. Hurricanes). Trajectory data are classified based on the layers as follows. *Preprocessing* is to improve quality of trajectory data and to partition trajectories into sub-trajectories for further processing. *Data Management* solves the problem of storing a huge amount of trajectory data in an efficient and scalable manner. *Query Processing* aims to retrieve appropriate data from underlying storage system efficiently. *Trajectory data mining tasks* are summarized and classified into several categories i.e. clustering, classification and knowledge discovery. *Privacy Protection* of users with privacy-preserving technique is essential problem throughout these four components above and combined with any component.

Zaiben Chen, Heng Tao Shen, and Xiaofang Zhou [10] proposed the discovery of finding the popular routes by investigating the Most Popular Routes (MPR) with analysis made by previous user. MPR will be beneficial to the user who is unfamiliar to the particular location. The huge collection of trajectory data gives the hint on how people usually travel between locations and discovers MPR. Initially the need is to retrieve a transfer network form the database to summarize user movement, transfer nodes and routes need to be established, finally combine the transfer edges to form the optimal route. *Coherence expanding algorithm* aim is to retrieve a transfer network from the raw trajectory which denotes all possible movements between the locations. *Absorbing markov chain* transfers the probability for each transfer nodes in the network. *Maximum Probability Product Algorithm* which is used to discover the MPR from a transfer network based on breath first manner.

III. METHODOLOGY

The destination prediction technique initially analyzed using the Most Popular Route based on the histories previous users between the locations. Location based services is related to the path discovery is extremely important to find the most suitable path that satisfy the predefined objective from source to destination. Using vector representation technique [7] the jerky movements are eliminated without any loss in information. The figure 1 shows the Representative Trajectory (ReTra) before and after merging.

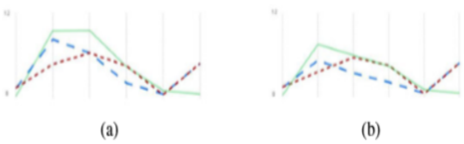

Figure 1 a) ReTra before merging [7]
b) ReTra after merging [7]

Next place prediction provides unprecedented opportunity to collect fine grained spatio-temporal data about the place user visit based on the Location Based Services. Based on the set of prediction features the exact place a user will visit next given historical data and current location such as historical visits or social ties.

Partition and summarization approach [5] has set of features defined for trajectory segments and finding the optimal partition and make the segments homogeneous as possible. Summarization provides the historical behaviour of the same route as compared with the common behaviour and generated by short text. The figure 2 [5] shows the trajectory data stored in real world and how it is stored in database.

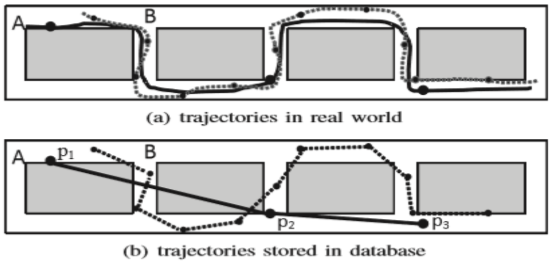

Figure 2 a) Trajectories in real world

b) Trajectories stored in database

IV CONCLUSION

Location prediction in trajectory data mining is the mapping between sources to destination using the GPS-equipped devices. Large scale trajectory data in discovering the location is a great challenge in the aggregate queries with stringent response-time constrain. The unwanted information need to be removed for proving the effectiveness and efficient way in storing the trajectory data in database. Next place prediction is based on the check-in of the user. The methodologies used above provide the feasible solution for mining the spatio-temporal trajectory data and can be used for several applications.

REFERENCES

1. Aiden Nibali and Zhen He, "Trajic: An effective compression system for trajectory data",IEEE Transaction on Knowledge and Data Engineering, vol. 27, no. 11, pp. 3138–3151, November 2015.
2. Anastasios Noulas, Salvatore Scellato, Neal Lathia, and Cecilia Mascolo, "Mining user mobility features for next place prediction in location-based services", in IEEE International Conference on Data Mining (ICDM), pp. 1038-043, December 2012.
3. Costas Panagiotakis, Nikos Pelekis et al., "Segmentation and sampling of moving object trajectories based on representativeness", IEEE Transaction on Knowledge and Data Engineering, vol. 24, pp.1328-343, July 2012.
4. Han Su, Kai Zheng et al."Calibrating trajectory data for spatio-temporal similarity analysis", VLDB Journal, vol. 24, no. 1, pp. 93–116, Feb 2015.
5. Han Su, Kai Zeng et al., "Making sense of trajectory data: A partition-and-Summarization approach", IEEE International Conference Data Engineering (ICDE), pp. 963-974, April 2015
6. Lyu-Chao, Xin-Hua Jiang ,Fu-Min Zou et.al,"A Method of Latent Semantic Information Mining for Trajectory Data", IEEE International Conference on Intelligent Information Hiding and Multimedia signal processing, pp.353-356, 2015.
7. Nikos Pelekis, Ioannis Kopanakis, Costas Panagiotakis, and Yannis Theodoridis,

"Unsupervised trajectory sampling" in Discovery Databases, Eur. Conf. (ECML PKDD), pp.17-33, Sep 2010.

8. Yanhua Li et al., "Sampling big trajectory data", ACM International Conference Information Knowledge and Management(CIKM), pp. 941-950, October 2015.
9. Zhenni Feng and Yanmin Zhu, "A Survey on Trajectory Data Mining: Techniques and Application", IEEE Access on Theoretical foundation for Big Data Application,Vol. 4, pp.2056-2067,April 2016.
10. Zaiben Chen, Heng Tao Shen, and Xiaofang Zhou, "Discovering popular routes from trajectories", ICDE Conference, pp. 900-911, April 2011.