

An Uncut Survey on Application and Techniques of Trajectory Data Mining

¹ K.Dharmarajan , ² A.Joynar Preethi and ³V.Jayalakshmi

¹. Department of Information Technology, School of Computing Sciences, VISTAS, India,

². Research Scholar, School of Computing Sciences, VISTAS, In

³. Department of Information Technology, School of Computing Sciences, VISTAS, India,

¹dharmak07@gmail.com , ²joynar.preethi@gmail.com, ³jayasekar1996@yahoo.co.in

ABSTRACT

The New generation of trajectory data is increased by fast advancement of accurate position acquisition and computation technologies. It tracks the traces of objects in motion such as people, vehicles and animals. Successive time stamped geographical location represents a typical trajectory. A Trajectory data mining is beneficial for a broad range of applications. The main contribution of this paper is to survey the application of trajectory data mining. For instance, Position forecasting, course finding, Behavioural analysis of movement, etc... Besides, this paper reviews an ample collection of existing trajectory data mining techniques and explores them in a concept of trajectory data mining. It also provides a structure which guides for designing a forthcoming trajectory data mining solutions.

KEYWORDS: Trajectory Data Mining, Applications of Trajectory Data Mining, Data mining techniques.

INTRODUCTION

Creating large-scale trajectory data of tracking traces of objects in motion is growing uncomplicated as multiple technologies provide positioning services, for instance GSM beacons, Global Position Systems (GPS), Radio Frequency Identification (RFID), position estimation of 802.11, infrared or ultrasonic systems, and smart phone sensors etc.

Trajectory is a sample location points that the object in motion passes as illustrated In Fig. 1. In geographical space, trace of object in motion is continual. A sequence of time stamped locations usually represent a spatial trajectory, for example in Fig 1: (P1, T1), (P2, T2)... (Pn, Tn). Trajectory sampling rate and duration is based on applications.

Smart phones, Vehicles equipped by GPS, Online check in data, Social network media, RFID readers, and Geo-tagged messages are the origin of Trajectory data. Accordingly, objects in motion can be a person, natural phenomena and vehicles

Trajectory data mining controls and enhances a broad range of application such as Position

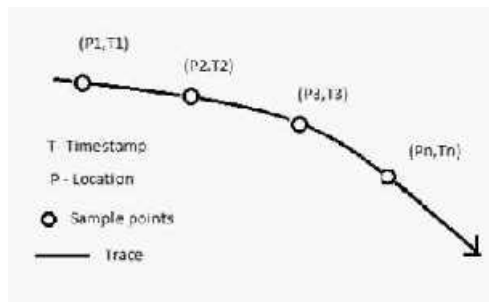
forecasting, course finding, Behavioural analysis of movement for group of objects in motion or individual, etc. Notably government agencies, common people and commercial organizations are benefited by these applications

Nevertheless, there is a challenge in managing, processing and mining trajectory data. For instance, diverse problems are taken here. In First, it is frivolous problem to stock an enormous size of trajectory data which are quickly accumulated. In Second, It is obdurate to describe a resemblance metric for differentiating trajectories. It is a primary feasibility in trajectory data mining. In third, It is hard to handle queries on the great amount of trajectory data in connection with complexity of time or space.

An immense collection of proposals have been put forward in order to tackle these problems. They are arranged in a manner corresponding to the trajectory data mining procedure. Besides, a structure is recommended for rationalizing these methods also it provides a uncut survey on trajectory data mining. Approximately, there are three levels in the structure. i.e., data acquiring, application and trajectory data mining techniques

Explicitly, five elements are listed below in the level of trajectory data mining.

1. Data Pre-processing: Trajectories are usually segmented, cleaned, sampled for representatives, or inferred from uncertain trajectories and calibrated In the level of data pre-processing
2. Data Organization: Trajectories are simplified before being collected. Moreover, efficient storage systems are built. Furthermore, appropriate index structures are also required in order to assist query processing.
3. Query computing: In order to fetch data, there are many queries have to be computed. e.g., Range queries, pattern queries, position-based queries, top-k queries, nearest



neighbour queries, aggregate queries and other application-specific queries. These queries are computed depends on index structure and an underlying storage system

4. Task of Trajectory data mining: Trajectory data mining tasks are reviewed and organized into several phases, i.e., clustering, pattern mining, knowledge discovery and classification.
5. Privacy preservation: In every steps of trajectory data mining, it is a crucial problem for privacy protection few examples are given to depict how to process trajectory data as well as to protect sensitive information of users.

Trajectory data mining

Data mining is a major step of a process, generally known as knowledge finding which extracts useful information from large datasets. Data mining methods and applications have been broadly surveyed in the general data mining domain. Some primary terminologies are described here, e.g., semantic trajectory, trajectory, semantic trajectory, path, road network

Trajectory data:

Depends upon the technology in which they are recorded, mobility data is available in various forms. Differentiated Global positioning system, Global system for mobile communications, and geo-social network based trajectory data. Radio frequency identification based and Wi-Fi based data.

Global positioning system based data is comprised of temporally ordered sequences of geographic coordinates recorded by a Global positioning system - enabled device carried by the object in motion. Global system for mobile communication based

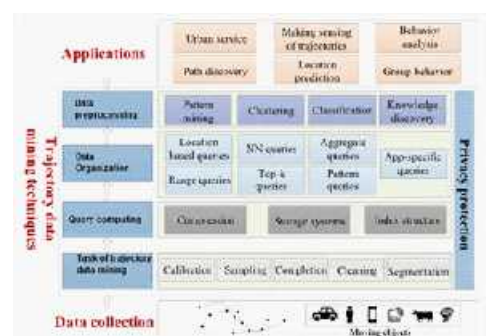
data is made up of temporally ordered sequences of identifiers of the cells in which the object in motion passes.

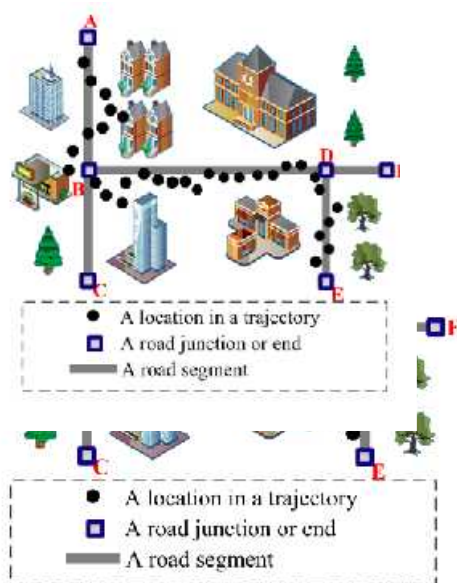
Geo-social network based data is content identified on Internet social media and to which geographic coordinates have been added. Radio frequency identification based data contain a sequence of identifiers of Radio frequency identification readers via the object in motion passed, while Wi-Fi based data contain a sequence of identifiers of access points that communicated with the objects in motion. Although the properties of these forms of data, example: their accuracy, show they are very different and they are used to address related application problems using similar mining methods. A trajectory of an object in motion is a discrete trace that the object in motion object travels in geographical space.

Usually, trajectory is a sequence of geo-locations with corresponding timestamps in spatio-temporal space, i.e., $T = (P_1, T_1), (P_2, T_2) \dots (P_n, T_n)$ where each member indicates an object in motion at position P_i at timestamp T_i . Moreover, members are sorted by timestamps, i.e., $T_j < T_k$ if $1 \leq j < k \leq n$.

An object in motion can be a vehicle, person, an animal, a mobile device, or even a phenomenon. A trajectory of an animal describes its trace generated by daily activities such as running. A trajectory of a person records one's trace for a period of time.

For instance, a trajectory of a person in a daytime may record his path to work in the morning and his path to home at night. A trajectory of a vehicle is recorded by a vehicle installed by Global positioning system device. At a fixed rate, it generally reports position of the vehicle, e.g., every second or every minute.





A special kind of trajectory data is Radio frequency identification data. There are two types of devices in RFID technology, i.e., tags by which radio signals emit with identification information and readers by which signals are detected from tags.

Mainly, an Object in motion is a tag device that is a good in a warehouse for example. A position is explained by identification of a reader by which signals are detected from that tag. Pertaining to the position of an object in motion are recorded by corresponding geographical areas of readers that detect its signals.

Trajectory data sampling rates changes greatly from data source to data source. Due to limitation in energy and storage, trajectories of different type of object in motion are sampled at various rates. Trajectory data of vehicles generally have larger sampling rates than those of mobile devices since vehicles can give enough battery and storage.

Specifically, a trajectory mentioned above can also be referred as raw trajectory or geographical trajectory. After that a special kind of trajectory data is introduced and that integrate geographical positions with semantic meaning.

Semantic Trajectory: A trajectory is referred as semantic trajectory when its positions are associated with semantic entities.

Semantic trajectories are often created by tagging a position point with a meaningful place in real world apart from numerical coordinates, such as check-in data. Furthermore, when considering trajectories created by vehicles, we often refer to a road network and paths in the road network.

When a geographical trajectory is associated with description text which expresses person's emotions and feeling it is also a type of semantic trajectory data.

Fig. 2 depicts an example of a road network and paths. Each position point in the collected trajectory is first matched to a specific road segment on the map. The path that the moving object travels is AB; BD; DE

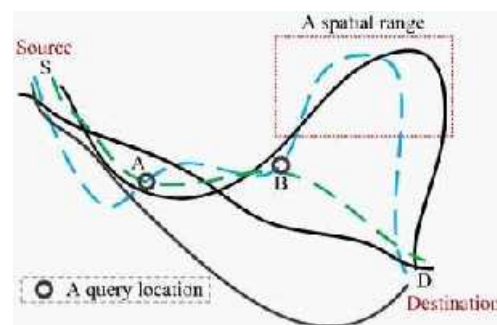
Road Network: A road network is a directed graph, G is equal to (V, E) , where E and V are a vertex set and an edge set, respectively. A vertex is a road junction or road end. An edge indicates a directed road segment.

Path: It indicates a sequence of edges in a road network, specifically, consecutive edges.

A STRUCTURE OF TRAJECTORY DATA MINING

Here, we propose a structure that summarizes a whole procedure of trajectory data mining as shown in Fig. 3. It is worth noting that not every step in the layer of trajectory data mining techniques is necessary and it is based on requirement of applications and collected data.

In First, trajectory data are created by different objects in motions and collected from various data sources. In the paper, we neglect details in data collection. Then, main part of trajectory mining



techniques are explained with five components, i.e., data pre-processing, data organization, query

computing, Task of trajectory data mining, and privacy preservation.

The phase of trajectory data mining techniques is classified as follows. Data Pre-processing tries to improve quality of trajectory data and to partition trajectories into sub-trajectories for additional processing. Data organization gives solution to the problem of storing a large amount of trajectory data in a scalable and efficient manner.

Query computing aims to get an appropriate data from the underlying storage system efficiently. The element of trajectory data mining tasks classifies little important kind of mining tasks. Protecting privacy of users with privacy-preserving techniques is an important problem throughout these four components above and thus it can be merged with any component.

The phase of applications and all elements of trajectory data mining techniques will be discussed successively in following sections.

DATA PREPROCESSING

Data Pre-processing is an initial step which carry out at the beginning and it aims at improving quality of trajectory data and generating sub-trajectories., it is presented here the common operations in pre-processing phase.

TRAJECTORY SEGMENTATION

In many application scenarios, a trajectory is partitioned into sub-trajectories [1], each of which is often called a segment, a partition or a frame. Generating sub-trajectories is rational because it corresponds to underlying structures in trajectory data, e.g., a path with multiple road segments

TRAJECTRY DATA CLEANING

Due to ambivalence of Radio frequency identification data, i.e., there existing no deterministic position given various readers identified an object, cleaning trajectory data targets to discard impossible positions or trajectories exploiting specific constraints, e.g., unreachability constraints maximum speed, [2]

TRAJECTORY CALIBRATION

Trajectories that refer discrete approximation of original routes with multiple sampling strategies and various sampling rates are heterogeneous. Heterogeneity has a negative effect on measurement of trajectory similarity, For instance: it is difficult to differentiate two trajectories derived with various sampling strategies by directly utilizing spatial proximity depends on resemblance measures like Euclidean distance.[3],[4]

TRAJECTORY COMPLETION

Owing to consideration of transmission and storage, trajectories are mostly possessed at relatively low sampling rates, only giving partial observations of actual routes. These trajectories are called uncertain trajectories [5], [6]

TRAJAECTORY SAMPLING

Trajectory sampling approaches targets to minimize a huge trajectory database appropriately, taking only the most representative samples of original trajectory database [7],[8]. The subset of samples should enclosed mobility patterns hidden in the original trajectory database.

DATA ORGANIZATION

Fundamental problem in trajectory data mining is how to store a large amount of trajectory data. In the section, we discuss the problem from two aspects.

TRAJECTORY DATA COMPRESSION

It is probably difficult to store or transmit a large amount of trajectory data created by location-acquisition techniques. As some position points in a trajectory are often redundant, trajectory compression algorithms [9]-[11] are promising to mitigate storage requirements and communication loads. A compression method is often made a trade-off between compression ratio and maximum error.

Usually, the larger the compression ratio, the lower the quality of compressed trajectory data. Typical methods are line generation and delta compression.

STORAGE SYSTEMS & INDEX STRUCTURES

A crucial problem by location-acquisition

technologies is how to store the large amount of trajectory data produced [12], [13].

In order to boost performance of query processing for trajectory data, a novel in-memory based trajectory storage system. To employ column-oriented storage, trajectories are partitioned into frames, and then frames are again compressed and well-structured to better support trajectory data mining.

A dynamic storage system is able to get all data in a particular region efficiently. Instead of simply indexing geo-spatial data, trajectories are sliced according to spatial-temporal regions and stores packed data in each region together. An index structure [13] is created to manage trajectory data especially for get tasks like k-NN queries.

QUERY PROCESSING

Recovering data from an underlying storage system is a critical operation. The objective of recovering is to find appropriate data efficiently.

POSITION-BASED QUERIES

A position-based query tries to find trajectories that are near to all query positions where the query is a tiny set of position with or without a specific order constraint. One typical application is route recommendation for a trip to various places.

For instance, in Fig. 4, a traveller at the source location S wants to find a correct path that passes query positions A and B to the destination explore this kind of problem and identify the k best-connected trajectories (k-BCT) that each of which connects all query positions.

TRAJECTORY RANGE QUERIES

Trajectory Range queries are necessary in a broad spectrum of trajectory data mining applications. A query that specifies a value to fall in a lower and upper boundary is regarded as range query, e.g., identifying all trajectories of a specific traveller between 3 am and 5 pm. In Fig 4, a spatial range query is perhaps to find all trajectories that pass a spatial region mentioned by a dashed rectangle.

TOP-k QUERIES

It focuses on identifying the k most similar trajectories for a said trajectory on uncertain

trajectories. The main key to a rational solution is to appropriately quantify the dissimilarity of two uncertain trajectories. A novel distance metric is proposed and a scalable index structure is further built to support trajectory data mining tasks

NEAREST NEIGHBOR (NN) QUERIES

Nearest neighbour find is another basic query in spatial-temporal trajectory data mining [14], [15] it address a problem of identifying the k nearest neighbours to a given trajectory in the trajectory database for a specific time interval. It extends nearest neighbours queries to a more realistic case, the probabilistic nearest neighbour queries, where trajectories are uncertain. Depends on representation of uncertain trajectories as stochastic processes, several kinds of probabilistic nearest neighbour queries with inputs of a given trajectory and a time interval are investigated.

TRAJECTORY AGGREGATE QUERIES

It introduces another type of queries called trajectory aggregate queries [16] whose querying results are not trajectories in a trajectory archive but an aggregated measurement. This type of query, as a kind of function, tries to retrieve statistics of trajectories that traverse a given region in a specific time interval. An example of an aggregate query for vehicular trajectory data is to retrieve average velocity passing a user-specific road segment

APPLICATION- MAKING SENSE OF TRAJECTORY DATA

Raw trajectory data in the form of course of geographical locations and timestamps fail to make sense to people without semantic description. There exist a great many studies [17], [18]-[21] to ease interference of raw trajectory data.

In contrast to semantic trajectory which cannot express movement properties of objects in motion, e.g., overspeed, stopover, Su et al. [17] Suggest a separation-and-summarization approach that automatically creates a short human-readable text to define a trajectory. The approach not only extends expressivity of traditional semantic trajectories but also neglect a challenging problem of processing, storage and transmission of large volume of semantic

trajectories. A raw trajectory data is partitioned in first according to behaviour of an object in motion and

then characteristics of each trajectory partition are summarized by short textual description. Besides, a proto system named STMaker [18] according to this idea is developed.

It is certainly worth noting that semantic meaning of position and short textual messages gathered by social media services provide an unprecedented opportunity to illuminate raw trajectory data. TOPTRAC aims to identify latent topic in trajectory data. Particularly, this method not only enlarges semantic regions with a coherent topic but also obtains mobility patterns of human beings between semantic regions. Likewise, Lu et al. [21] use a clustering-based approach to fine semantic regions.

A lot of emerging positions-aware applications need a semantic notation of a position point, e.g., ``work, ``home, in replacement of longitude and latitude coordinates. Lv et al. [22] desire a approach of automatically finding personal semantic places (i.e., both a physical location and semantic meaning of the position).

XI. CONCLUSION

Individual persons are benefited by Trajectory data mining. Also the person can understand their movement behaviour better through examining historical trajectories. Furthermore, trajectory data mining gives plenty of comfort to the public, For example. En-route Guidance, Real-time traffic information by transport agencies. Anyhow, people be affected from privacy breaches if their trajectories are inappropriately gathered and used.

Furthermore, citizens are generally issued by commercial advertisements those are possibly pushed in the name of personalized services.

Trajectory data mining assist to mitigate the cost of supervision for the government and organization. In urban areas, trajectory data mining from vehicle trajectories gives an efficient and scalable method to monitor traffic condition of the whole city. One more example is to record illegal or irregular behaviour which perhaps priceless to ascertain responsibilities later. For instance, rapidity in moving can be deduced

from trajectories. This affirmation is precious predominantly in roads without wayside cameras. Comparably, private enterprises await to cut down their costs in lieu of trajectory data mining. For instance, RFID data, as a special kind of trajectories, significantly help to organize artefact stocks.

Position acquisition technologies produce large amount of trajectory data. Trajectory data which track path of object in motion objects is typically represented by a sequence of time stamped geographical positions. A huge amount of applications are generated based on trajectory data mining. The survey inspects a great collection of existing studies in the recommended structure of trajectory data mining.

References

- [1] N. Pelekis, I. Kopanakis, C. Panagiotakis, and Y. Theodoridis, ``Unsupervised trajectory sampling,’’ in Proc. Mach. Learn. Knowl. Discovery Databases, Eur. Conf. (ECML PKDD), Barcelona, Spain, Sep. 2010, pp. 17 33.
- [2] B. Fazzinga, S. Flesca, F. Furfaro, and F. Parisi, ``Cleaning trajectory data of RFID-monitored objects through conditioning under integrity constraints,’’ in Proc. 17th Int. Conf. Extending Database Technol. (EDBT), Athens, Greece, Mar. 2014, pp. 379 390.
- [3] H. Su, K. Zheng, H. Wang, J. Huang, and X. Zhou, ``Calibrating trajectory data for similarity-based analysis,’’ in Proc. ACM SIGMOD Int. Conf. Manage. Data (SIGMOD), New York, NY, USA, Jun. 2013, pp. 833 844.
- [4] H. Su, K. Zheng, J. Huang, H. Wang, and X. Zhou, ``Calibrating trajectory data for spatio-temporal similarity analysis,’’ VLDB J., vol. 24, no. 1, pp. 93 116, Feb. 2015.
- [5] M. Li, A. Ahmed, and A. J. Smola, ``Inferring movement trajectories from GPS snippets,’’ in Proc. 8th ACM Int. Conf. Web Search Data Mining (WSDM), Shanghai, China, Feb. 2015, pp. 325 334.
- [6] M. Chiang, Y.-H. Lin, W.-C. Peng, and P. S. Yu, ``Inferring distant-time location in low-sampling-rate

- trajectories," in Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD), Chicago, IL, USA, Aug. 2013, pp. 1454-1457.
- [7] C. Panagiotakis, N. Pelekis, I. Kopanakis, E. Ramasso, and Y. Theodoridis, "Segmentation and sampling of moving object trajectories based on representativeness," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 7, pp. 1328-1343, Jul. 2012.
- [8] Y. Li et al., "Sampling big trajectory data," in Proc. 24th ACM Int. Conf. Inf. Knowl. Manage. (CIKM), Melbourne, VIC, Australia, Oct. 2015, pp. 941-950.
- [9] A. Nibali and Z. He, "Trajic: An effective compression system for trajectory data," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 11, pp. 3138-3151, Nov. 2015.
- [10] R. Song, W. Sun, B. Zheng, and Y. Zheng, "PRESS: A novel framework of trajectory compression in road networks," in Proc. PVLDB, 2014, vol. 7, no. 9, pp. 661-672.
- [11] J. Liu, K. Zhao, P. Sommer, S. Shang, B. Kusy, and R. Jurda, "Bounded quadrant system: Error-bounded trajectory compression on the go," in Proc. 31st IEEE Int. Conf. Data Eng. (ICDE), Seoul, South Korea, Apr. 2015, pp. 987-998.
- [12] P. Cudré-Mauroux, E. Wu, and S. Madden, "Trajstore: An adaptive storage system for very large trajectory data sets," in Proc. 26th Int. Conf. Data Eng. (ICDE), Mar. 2010, Long Beach, CA, USA, 2010, pp. 109-120.
- [13] S. Ranu, D. P. A. D. Telang, P. Deshpande, and S. Raghavan, "Indexing and matching trajectories under inconsistent sampling rates," in Proc. 31st IEEE Int. Conf. Data Eng. (ICDE), Seoul, South Korea, Apr. 2015, pp. 999-1010.
- [14] R. H. Güting, T. Behr, and J. Xu, "Efficient k-nearest neighbor search on moving object trajectories," *VLDB J.*, vol. 19, no. 5, pp. 687-714, Oct. 2010.
- [15] J. Niedermayer, A. Zü e, T. Emrich, M. Renz, N. Mamoulis, L. Chen, and H.-P. Kriegel, "Probabilistic nearest neighbor queries on uncertain moving object trajectories," in Proc. PVLDB, 2013, vol. 7, no. 3, pp. 205-216.
- [16] Y. Li et al., "Sampling big trajectory data," in Proc. 24th ACM Int. Conf. Inf. Knowl. Manage. (CIKM), Melbourne, VIC, Australia, Oct. 2015, pp. 941-950.
- [17] H. Su et al., "Making sense of trajectory data: A partition-and-summarization approach," in Proc. 31st IEEE Int. Conf. Data Eng. (ICDE), Seoul, South Korea, Apr. 2015, pp. 963-974.
- [18] H. Su, K. Zheng, K. Zheng, J. Huang, and X. Zhou, "STMaker: A system to make sense of trajectory data," in Proc. PVLDB, 2014, vol. 7, no. 13, pp. 1701-1704.
- [19] Y. Kim, J. Han, and C. Yuan, "TOPTRAC: Topical trajectory pattern mining," in Proc. 21st ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Sydney, NSW, Australia, Aug. 2015, pp. 587-596.
- [20] M. Lv, L. Chen, and G. Chen, "Discovering personally semantic places from GPS trajectories," in Proc. 21st ACM Int. Conf. Inf. Knowl. Manage. (CIKM), Maui, HI, USA, Oct./Nov. 2012, pp. 1552-1556.
- [21] C. Lu, P.-R. Lei, W.-C. Peng, and I.-J. Su, "A framework of mining semantic regions from trajectories," in Proc. 16th Int. Conf. Database Syst. Adv. Appl. (DASFAA), Hong Kong, Apr. 2011, pp. 193-207.
- [22] Dharmarajan, K., and M. A. Dorairangaswamy. "Discovering Student E-Learning Preferred Navigation Paths Using Selection Page and Time Preference Algorithm." *International Journal of Emerging Technologies in Learning (iJET)* 12.10 (2017): 202-211.

