ISSN XXXX XXXX © 2018 IJESC

Research Article Volume 8 Issue No.10

Standardization of Pre-processed Trajectory Data Representation

AdleneEbenezer[1], Vikas Malviya[2], Vikram Sinha[3], Manjeet[4]
Assistant Professor[1], Computer Science and Engineering[2, 3, 4]
SRM Institute of Science and Technology, India

Abstract
Trajectory data Mining is used in various applications such as traffic controlling, migration problems, tourism, satellite positioning data etc. The increasing use of Trajectory Data in different applications makes it very important that the data should be handled properly and it should be easily understandable by different applications without any modifications. But the existing systems working on the trajectory data have no standard or pre-defined method for handling the raw data before mining from it. The absence of standard method causes problems such as: Un-portable data i.e. data being used for one application cannot be used for other applications without doing modifications. These modifications are not error proof and may cause loss of data causing inefficiency. This project provides (1)a set of trajectory pre-processing techniques that can be used as a standard raw data pre-processing set and (2)a data model that can be used as a standard structure for all trajectory related problems to achieve better efficiency during actual data mining.

Keyword - Trajectory Data Pre-processing, Standardization, Semantics

1. Introduction

The increasing number of location aware devices that make use of, as well as produce trajectory data at rates that require huge amount of storage, proper classification and efficient data retrieval processes make the field of trajectory data mining a very important area of data analysis. Also, the absence of a standard trajectory data model that takes into consideration the position, time and a semantic location calls for a proper address to this area.

Yu Zheng defines spatial trajectories as, [1] a trace generated by a moving object in geographical spaces, usually represented by a series of chronologically ordered points, for example, p1 → p2 →⋯→ pn, where each point consists of a geospatial coordinate set and a time stamp such as p= (x, y,t).

Jean Damascène Mazimpaka and Sabine Timpf represent spatial trajectories as, [3]A trajectory can be generally formally represented as: T=(p1...pn)where pk=(id(k), loc(k),t(k), A(k)) is the kth position, id(k) is the position identifier, lock is the spatial location of the position, t(k) is the time at which the position was recorded, and A(k) is a possibly empty list of additional descriptive data (e.g., direction, occupancy status, etc.). The spatial location of the position may be represented in different ways depending on the recording technology.

Qiang Ma, Bin Yang, Weining Qian, Aoying Zhou define spatial trajectory as, [4] a polyline in three-dimensional space, where two dimensions refer to space and the third dimension refers to time. It is represented as a sequence of position points Tr(P1, P2, ..., Pn); where each position point Pi is in form of (x,y,t), where x and y represent the coordinates of the object in a given spatial coordinate system, and t indicates the corresponding timestamp of the report of the position. Each line segment in a trajectory is indicated as Si(Pi, Pi+1), where Si is the line segment identifier, Pi and Pi+1 indicate two consecutive reported positions.

Similarly, there are many more that worked withtrajectory data. It is evident from the previously mentioned definitions that there is an absence of a standard for representing spatial trajectory data.

The absence of a standard produces a set of problems which arises every time a work is to be done on this kind of data. Some of which are mentioned below:

1. Lack of a standard representation makes the data difficult to work with as even the pre-processed data sometime doesn't fall into a desired pattern required by the individual.
2. The individual is required to device methods to convert the raw data to pre-processed data through methods which might not be efficient.
3. Wastage of time in devising the above-mentionedmethod to obtain pre-processed data.

This paper is an approach to provide a solution to this problem by providing a standard notation for representing the trajectory data by pre-processing it with different methods which are proposed in [1] and [2].

The paper is organized in the following manner:

1. Introduction
2. Trajectory Pre-processing techniques
   - 2.1 Noise Filtering
   - 2.2 Stay Point Detection
   - 2.3 Trajectory Compression
   - 2.4 Segmentation
   - 2.5 Map Matching and Semantics
3. Proposed Data Model
4. Conclusion and Future Works

References

2. Trajectory Pre-processing techniques

This section briefly explains a few trajectory pre-processing techniques that will be used for producing the required standard data model. For more information of the processes, refer to [1].

2.1 Noise Filtering

Noise filtering is the process of removal of erroneous trajectory points that do not follow certain conditions. Yu Zheng in [1] uses three different methods to produce noise filtered data. Out of the three, the Heuristics-based Outlier

detection is the method used in this paper as it also helps eliminate the pre-processing technique Out-lier Detection and saves time.

Heuristics-based Outlier Detection.

The heuristics-based outlier detection as the name suggests uses Outlier Detection algorithms to find the trajectory points that might be erroneous.

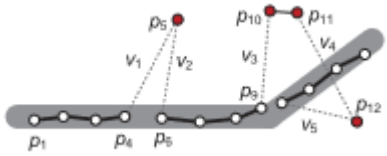

Fig 1. Noise points in a trajectory

This method uses the timestamp attached with each point and the time interval between adjacent points to calculate the speed of the object. The time intervals that have a larger speed then a threshold defined for that object are eliminated.

For example, in the Fig 1., the distance between points p4-p5, p5-p6 etc. is too much and the speed of the object will have to be much greater than the threshold to be able to reach the point that far in the given time interval. Hence the point p5 will be eliminated. This method has also been used in T-Drive[Yuan et al. 2010a, 2011a, 2013a] and GeoLife [Zheng et al. 2009a; Zheng et al. 2010].

2.2 Stay Point Detection

Trajectory points can be classified into moves and stops (stay points). These stops denote that the object stayed at a place for a longer duration of time. This process is also important as it is used for adding semantics to the raw data which is one of the important components of the proposed data model.These stay points can be used to convert the whole data from a timestamped trajectory to a sequence of meaningful places.

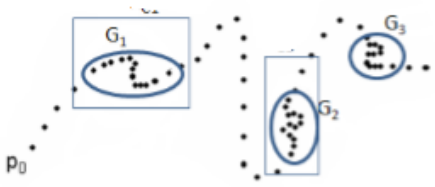

Fig 2. Stay Points in trajectory

The time interval between each point is usually constant and also the maximum and minimum speed of the object under normal conditions are well defined. Hence if the density of trajectory points in a certain area is large enough, it can be classified as a stay point. For this, the distance between adjacent points is compared to a defined minimum distance. If it is smaller than the defined distance, then the points can be classified as a stay point and a different notation can be used to represent these points. Grouping of these stay points depends on the needs of the application.

As shown in Fig 2., the density of points in the regions G1, G2 are greater than the defined density. Hence, they can be declared as stay points.

2.3 Trajectory Compression

Trajectory data is recorded in constant time intervals of seconds. But most applications don't need such precision in the data and the overhead cost and space for storing this data. Trajectory compression techniques play an important role in reducing the size of these trajectories by removing points that do not compromise the efficiency of the trajectory data. Many different methods are proposed in [1], of which Douglas-Peucker algorithm is used here.

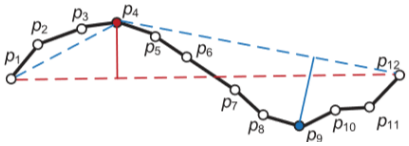

Fig 3. Trajectory Compression

Douglas-Peucker Algorithm.

In this algorithm, a virtual line segment is drawn between the first and the last point. Then the distance between all other points (between the end points of the line segment) is measured from the line segment. The distance of the point, that has the maximum distance from the line segment is then compared with a predefined error checkdistance. If the distance of the point from the line segment is greater than the error check distance, than the new point is considered as an end point of two new line segment drawn between the first point, new obtained point and the last point.

As shown in Fig 3., p1 and p12 are chosen as first and last point and last point respectively. P4 lies farthest from the line segment between these points. Its distance is compared with the error check distance and if found greater, p4 is considered as a new point. The same process is done again with line segment p1-p4 and p4-p12. After doing this for a few times, the elimination of points takes place and a trajectory is obtained with fewer points and hence data.

2.4 Segmentation

Segmentation is the process of dividing the trajectory into smaller sub-trajectories to obtain richer data and to reduce computational complexities.

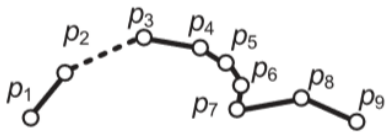

Fig 4. Segmentation of Trajectory Data

Time Interval-based Segmentation.

A simple method which segments the trajectories when the time interval between two adjacent points in a trajectory is greater than a predefined threshold.

As shown in Fig 4., the time interval between the points p2 and p3 is large enough to segment the whole trajectory into two sub-trajectories p1-p2 and p3-p9.

2.5 Map Matching and Semantics

Map matching is the process of converting a set of latitudes and longitudes coordinates into a sequence of roads. Only in this paper, the coordinates of stay points are also matched to provide semantics to the stay points by mapping them to places of interest to the moving object.

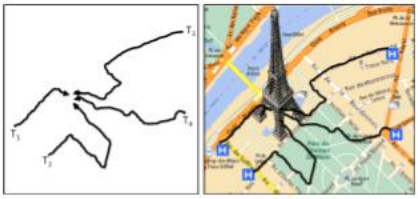

Fig 5. Map Matching and Semantics

All the processes described till now were in one way or the other adding semantics to the raw data which is the third parameter of the data model that is proposed below.
Semantics has been considered as an important data for representation of the trajectory data in this paper as it provides better understandability to humans.

3. Proposed Data Model

The proposed data model is a group of timestamped points with places they represent attached to the point.
(P0, T0, S0), (P1, T1, S1), …, (Pn, Tn, Sn) is the structure of the proposed trajectory data model where:
Pn is a trajectory points with n>=0,
Tn is the timestamp attached to Pn with n>=0,
Sn is a word or null value representing the place the point is mapped to. Null values can be used to address the points that do not correspond to any position of interest or points that are mapped to roads.
The addition of the third variable to the data model is considered necessary for the ease of understandability of these points and to highlight the points that are important to human understanding of the trajectory.

4. Conclusions and Future Works

In this paper, an absence of a standard way of representing pre-processed trajectory data has been shown. It then proposesa brief description of the processes of trajectory data pre-processing (with an algorithm of each process) which will help achieve the proposed data model which in subsequently described.
The future works will be based on making of an application that will take the raw data and will produce the pre-processed data in the format of the data model proposed. This will save a lot of time of the other applications dealing with trajectory data and the problems associated with this kind of data.

References

[1] Yu Zheng. Trajectory Data Mining: An Overview, Microsoft Research. ACM Transactions on Intelligent Systems and Technology, Vol. 6, No. 3, Article 29, Publication date: May 2015

[2] LuisOtavioAlvares, Gabriel Oliveira, Vania Bogorny. A Framework for Trajectory Data Pre-processing for Data Mining Instituto de Informatica – Universidade Federal do Rio Grande do SulPorto Alegre – Brazil

[3] JeanDamascèneMazimpaka and Sabine Timpf. Trajectory data mining: A review of methods and applications Department of Geography, University of Augsburg, Germany, Accepted: September 6, 2016

[4] Qiang Ma, Bin Yang, Weining Qian, Aoying Zhou. Query Processing of Massive Trajectory Data based on MapReduce

[5] Zhenni Feng,Yanmin Zhu. A Survey on Trajectory Data Mining: Techniques and Applications.IEEEAccess, accepted March 27, 2016, date of publication April 13, 2016.Department of Computer Science and Engineering, Jiao Tong University, KeyLaboratory of Scalable Computing and Systems,Shanghai.

[6] Muhammad Muzammal, Moneeb Gohar, Arif Ur Rahman, Qiang Qu1, Awais Ahmad, Gwanggil Jeon. Trajectory Mining Using Uncertain Sensor Data.IEEE Access, Accepted November 13, 2017, date of publication December 15, 2017.Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Department of Computer Science, Bahria University, Islamabad, Faculty of Computer Science, Free University of Bolzano,Laboratory of Machine Perception, Peking University, Beijing, Department of Information and Communication Engineering, Yeungnam University, Gyeongbuk, Department of Embedded Systems Engineering, Incheon National University, Incheon.