CLASSIFICATION ON MOVING OBJECT TRAJECTORIES

Sonal Athavale
Research Scholar
Rungta College of Engineering and Technology, Bhilai
Neelabh Sao
Reader
Rungta College of Engineering and Technology, Bhilai

Abstract

Trajectory means a path followed by a vehicles or object moving in space and time. Each point in this path represents one position in space and one instant in time. In our everyday action the way we live and move, leave digital traces in the information system. The potential value of these traces in recording human activities with reference to geography places is known as movement data or trajectory. These are used by traffic engineers and city managers. Trajectory of a moving object is a sequence of consecutive locations of the object in a multidimensional space. These are used by traffic engineers and city managers for observation of a movement patterns and to get behaviour of vehicles. As the no. of moving vehicles increase rapidly everyday the need for analysis, modelling and Pre-processing of traffic data is vital. In this project firstly we are using incremental DBSCAN Clustering technique after that we are classifying our data set to generate useful patterns using C 4.5 Algorithms.

Keywords: Trajectory, C 4.5 Algorithm, Moving Object

I. Introduction

Data mining can be performed on data represented in quantitative, textual, or multimedia forms. Data mining applications can use different types of parameters to examine the data. They include association, sequence or path analysis, classification, clustering and forecasting. Data mining has become increasingly common in both the public and private sectors “Ref [1]”. The comprehension of phenomena related to movement not only of people and vehicles but also of animals and other moving objects has always been a key issue in many areas of scientific investigation or social analysis. Spatio – temporal clustering is a process of grouping objects based on their spatial and temporal similarity. It is also known as trajectory or mobility data. Typically, Trajectory data are obtained from mobile devices that capture the position of an object at specific time intervals. It is relatively new subfield of data mining which gained high popularity especially in geographic information sciences.

II. Dbscan Clustering

Density Based Spatial Clustering of Applications with Noise (DBSCAN) has the ability to produce arbitrary shape of clusters. Clusters are identified by looking at the density of points. The key idea of density-based clustering is that for each object of a cluster the neighborhood of a given radius (ε) has to contain at least a minimum number of objects (MinPts). DBSCAN requires two parameters: ε (eps) and the minimum number of points required to form a cluster (Minpts). It starts with an arbitrary starting point that has not been visited. This point's ε-neighborhood is retrieved, and if it contains sufficiently many points, a cluster is started. Otherwise, the point is labeled as noise.

III. Incremental Dbscan

In incremental approach the DBSCAN algorithm is applied to a dynamic database where the data may be frequently updated. After insertion and deletion to the dynamic database the clustering discovered by DBSCAN has to be updated. And this approach measure the new cluster by directly compute the new data entering into the existing clusters instead of rerunning the algorithm. It finally discovers new updated clusters and outliers as well. Thus it describes at what percent of delta change in the original database the actual and incremental DBSCAN algorithms behave like same. DBSCAN is widely used in those situations where large multidimensional databases are maintained such as Data Warehouse.

IV. Problem Identification

As the number of moving vehicles increase rapidly everyday the need for analysis, modeling and preprocessing of traffic data is vital. When we think of moving object or trajectory data that represent traffic situated in some city or province, obvious task we would like to perform concerning everyday phenomena include detecting traffic jams, predicting traffic jams and discovering relations between traffic jams. We will also try to reduce the chances of occurring accidents [7][8].
V. Solution Approach

In this project firstly we are clustering our data set by using incremental DBSCAN clustering algorithm to find that how many vehicles are present at any particular locations at any particular time.

After that we are using C 4.5 Classification algorithm to generate useful pattern. So after clustering we are going to classify our data set which will help us to generate useful patterns. By Classifying our dataset we propose 3 categories of moving data.

1. **Directed/Precise**, means people are having fixed source and destinations, for ex. People goes to office daily follow same route daily comes under this category.

2. **Finding/Behaviour**, means their destination is not fixed but somewhat related, for ex. If any person goes to market to buy something so his destination is not fixed he will go to different shops but somehow will be related comes under this category.

3. **Roaming/Vague**, means some people just roam here and there on vehicles and their path is not decided comes under this category.

After that we will compare the result of clustering and classification and try to generate useful patterns and check the accuracy of both. All the classification will be done using C 4.5 classification algorithm which generates decision tree.

A. C 4.5 Classification Algorithm

C 4.5 is an algorithm used to generate a decision tree. C 4.5 is an extension of ID 3 Algorithm. The Decision tree generated by C 4.5 can be used for classification and for this reason C 4.5 is often referred to as a Statistical Classifier. C 4.5 builds decision tree from a set of training data in the same way as ID3 using the concept of information entropy.

At each node of the tree C4.5 chooses one attribute of the data that most effectively splits its set of samples into subsets enriched in one class or the other. Its criterion is the normalized information gain that results from choosing an attribute for splitting the data. The attribute with the highest normalized information gain is choosen to make the decision. The C 4.5 algorithm then recourses on the smaller sublists.

C 4.5 Algorithm for building Decision Tree is:-

1. Check for base cases

2. For each attribute a
   I. Find the normalized information gain from splitting on a
   3. Let a_best be the attribute with the highest normalized information gain
   4. Create a decision node that splits on a_best
   5. Recurse on the sublists obtained by splitting on a_best and add those nodes as children of node.

VI. Conclusion

In this our work focuses on management and maintenance of traffic system or for diverting traffic under certain emergency circumstances.

For this firstly we are clustering our data set using incremental DBSCAN algorithm and after that we are going to perform classification on that. After performing classification on that data set we will compare the performance of clustering and classification.

VIII. References


[7] Applications Volume-9 –No. 5. Jae-Gil Lee, Jiawei Han & Kyu-Young Whang” Trajectory Clustering: A Partition-
and-Group Framework in University of Illinois at Urbana-Champaign KAIST


**Biographies**

1. SONAL ATHAVALE Completed her B.E. Degree in Computer Science & Engineering Branch from Chouksey Engineering Coll. Bilaspur C.G.,India in 2009 and pursuing M. Tech in Computer Technology Branch from Rungta College of Engineering and Technology Bhilai. Sonal Athavale may be reached at sonal.athavale@gmail.com

2. Mr. NEELABH SAO Completed his B.E. Degree in Computer Science & Engineering Branch from Rungta College of Engineering And Technology, Bhilai C.G., India in 2003 and Completed his M.Tech Degree in Computer Science & Engineering Branch from Rungta College of Engineering & Technology, Bhilai, C.G., India in 2009. Neelabh Sao may be reached at neelabhsao@gmail.com