Journal of Traffic and Transportation Engineering(English Edition)
2014,1(6):393-403

Finding frequent trajectories by clustering and sequential pattern mining

Arthur A. Shaw*, N. P. Gopalan

Department of Computer Applications, National Institute of Technology, Tiruchirappalli, Tami Nadu, India

Abstract: Data mining is a powerful emerging technology that helps to extract hidden information from a huge volume of historical data. This paper is concerned with finding the frequent trajectories of moving objects in spatio-temporal data by a novel method adopting the concepts of clustering and sequential pattern mining. The algorithms used logically split the trajectory span area into clusters and then apply the k-means algorithm over this clusters until the squared error minimizes. The new method applies the threshold to obtain active clusters and arranges them in descending order based on number of trajectories passing through. From these active clusters, inter cluster patterns are found by a sequential pattern mining technique. The process is repeated until all the active clusters are linked. The clusters thus linked in sequence are the frequent trajectories. A set of experiments conducted using real datasets shows that the proposed method is relatively five times better than the existing ones. A comparison is made with the results of other algorithms and their variation is analyzed by statistical methods. Further, tests of significance are conducted with ANOVA to find the efficient threshold value for the optimum plot of frequent trajectories. The results are analyzed and found to be superior than the existing ones. This approach may be of relevance in finding alternate paths in busy networks (congestion control), finding the frequent paths of migratory birds, or even to predict the next level of pattern characteristics in case of time series data with minor alterations and finding the frequent path of balls in certain games.

Key words: data mining; frequent trajectory; clustering; sequential pattern mining; statistical method

1 Introduction

Principles of data mining may be applied to a large volume of data to extract some new non-trivial information. With the pervasiveness of cutting edge technologies, there is an increased availability of huge volumes of data pertaining to trajectories of moving objects collected from communication equipments like, GPS, GSM networks etc., and the logs they had created. The frequent trajectory (FT) patterns from these data may provide the information on the overall cumulative behavior of a population of moving

* **Corresponding author**: Arthur A. Shaw, M. Tech, Research Scholar.
E-mail: Arthurshaw123@gmail.com.

objects like their mobility pattern. This might help finding alternate paths in congestion control. In games, the FT of ball struck by a player may be found and consequently the opponents can play defensively and increase their chance of winning. In case of migratory birds, using the information obtained from their FT, air traffic can be channelized to avoid bird hits on flights. Hence finding FT assumes significance and it can be applied with some additional statistical concepts to predict the next level of information in the time series domain.

Recently substantial researches were performed in this area to find the frequent trajectory patterns (Duda et al. 2001; Laube et al. 2005; Alvares et al. 2007; Gudmundsson et al. 2007; Lee et al. 2009; Shaw 2010a-d). This research paper highlights the overview of existing frequent trajectory pattern methods, its extensions and applications. Consider a large number of trajectories passing over an area of interest. This area may be divided into clusters of various sizes (Cf. Section 4.7). A cluster count is obtained from the ratio of the number of coordinates of the trajectories passing through it to the total number of coordinates of all trajectories. The cluster count of the cluster is greater than the minimum support threshold value, such cluster may be called as the active cluster. They are sorted in descending order of their count. Such clusters are linked together by a sequential pattern mining technique and the FT of the moving objects can be found out. The method is discussed in Section 3.1, and is proven to be relatively the best among the existing ones in this category with respect to time and space complexities. As the algorithm is implemented in Java with size of required memory less than one MB, it may be easily implemented across various platforms.

With the rich content of literature on this theme, this paper is organized into the following three themes in successive sections: related work; mining frequent trajectories of moving objects; results and conclusions.

2 Related work

Trajectories left behind by moving objects are accumulated to form a new kind of data. Such data are normally available in the form of spatial or spatio-temporal coordinates in two or three dimensions which may have very little or no semantics. From these data no useful information can be gathered by commuters that wish to explore the possible routes for a place. This difficulty can be overcome by applying the data mining principles and gathering useful information about the moving object's behavior or patterns. It is a complex process to extracting information from the trajectory database (Duda et al. 2001; Brakatsoulas et al. 2004; Laube et al. 2005; Mouza and Rigaux 2005; Alvares et al. 2007; Cao et al. 2007; Giannotti et al. 2007; Gudmundsson et al. 2007; Spaccapietra et al. 2008; Lee et al. 2009; Shaw 2011). Some of the developments over the last two decades focused on their spatial and temporal properties. Mouza and Rigaux (2005) proposed a data model for tracking mobile objects by partitioning the spatio-temporal data into spatial and temporal components and retrieve objects that match mobility patterns by framing deterministic queries. Semantic data models were proposed for queries on moving objects (Bishop 2009). But usually, querying is not the objective of data mining. In certain cases researchers considered semantics and background or environment geographic information of trajectories (Cao et al. 2007; Spaccapietra et al. 2008; Lee et al. 2009). As the moving objects follow some patterns over a period of time, their paths can be mined from their spatial attributes. Spaccapietra et al. (2008) proposed some basic definitions and semantics for trajectories which are focused on finding the frequent trajectories in the object space. Cao et al. (2007) defined the problem of mining periodic pat-terns in spatio-temporal data, but for finding the patterns by clustering the time component was unutilized. Distance travelled by an object is considered for finding the frequent patterns in (Laube et al. 2005). Brakatsoulas et al. (2004) and Mouza and Rigaux (2005) considered the semantics and geographic information behind trajectories for finding trajectory patterns. Trajectories can be classified according to their spatial coordinates, time component, velocity and direction of movement. A trajectory may fall in a sequence of clusters (Chen et al. 2009). The K-means algorithm has been widely ap-