Mining Moving Object, Trajectory and Traffic Data

Jiawei Han, Zhenhui Li, and Lu An Tang

Department of Computer Science
University of Illinois at Urbana-Champaign, USA
{hanj,zli28,tang18}@uiuc.edu

With the wide availability of satellite, RFID, GPS, sensor, wireless, and video technologies, moving-object data have been collected in massive scale and are becoming increasingly rich, complex, and ubiquitous. There is an imminent need for scalable and flexible data analysis over moving-object information; and thus mining moving-object data has become one of major challenges in data mining. There have been considerable research efforts on data mining for moving object, trajectory, and traffic data sets. However, there has been few systematic tutorial on knowledge discovery from such moving-object data sets. This tutorial presents a comprehensive, organized, and state-of-the-art survey on methodologies and algorithms on analyzing different kinds of moving-object data sets, with an emphasis on several important mining tasks: *pattern-mining, clustering, classification, outlier analysis*, and *multidimensional analysis*. Besides a thorough survey of the recent research work on this topic, we also show how real-world applications can benefit from data mining of moving object, trajectory, and traffic data sets. The tutorial consists of three parts: (1) moving object pattern mining, (2) trajectory data mining, and (3) traffic data mining.

In the first part, *moving object pattern mining*, we introduce different pattern mining algorithms for various moving object patterns. *Frequent pattern* is one of the most basic patterns that detects frequently visited routes, such as "Railway Station → Time Square → Central Park" for New York city travelers. The challenge of frequent pattern lies in the approximation of locations and transition time. *Periodic pattern mining* is another interesting topic since periodicity is an intrinsic nature of moving objects. For example, people have weekly working pattern and animals have yearly migration pattern. But how to detect periodicity for moving objects in 2-dimensional space remains a difficult problem. *Moving object grouping pattern* discovers the social behaviors of moving objects in groups. Research have been conducted in different definitions, such as *moving cluster*, *flock* and *convoy*. But strict temporal and spatial constraint may result in failure of finding meaningful patterns. Thus, a concept called *swarm* is further proposed to suit for more realistic cases. Other interesting patterns including *leadership*, *following*, and *meeting* are studied as well. Leadership and following patterns discovers a small number of objects (*e.g.*, suspect, wolf) that follow one or a set of given moving objects (*e.g.*, people, sheep). Meeting could describe the movement that suspects meet to plot an attack or animals meet together for the same food resources.

In the second part, *trajectory mining*, we focus more on trajectory clustering, classification and outlier detection. For *trajectory clustering*, many high

H. Kitagawa et al. (Eds.): DASFAA 2010, Part II, LNCS 5982, pp. 485–486, 2010.

dimensional data clustering methods can be easily adapted if we treat each timestamp as one dimension. Studies related to different distance measures in high dimensional space and efficient computation of these distance functions haven been conducted. Probabilistic methods are also proposed by modeling a set of trajectories as individual sequences of points generated from a finite mixture model. While these methods cluster trajectories as a whole, they ignore sub-trajectory clusters. A partition-and-group framework is proposed to solve this problem. It first partitions a trajectory into multiple line segments and then cluster the line segments based on density. Outlier could be a natural byproduct of clustering result. The objects that are distant from any cluster can be considered as outliers. Recently, methods specifically designed for more complex cases are developed, such as integration of multi-dimensional information and partial trajectory outlier detection. While there are many clustering and outlier detection methods, few *classification method* has been developed for moving objects. A related area could be time series classification, in which 1-Nearest-Neighbor has been the most popular method and a shapelet-based classification method has recently shown to be effective. Time series studies mainly deal with 1-dimensional data whereas moving object classification has to face more complicated 2-dimensional spatial data. Recently, a trajectory classification method based on regions and trajectory clusters has shown satisfactory result. This method extracts discriminative regions and trajectory clusters for some class as classification features.

Finally, we introduce some state-of-the-art traffic data mining methods. Traffic data, different from free space movement, is confined to road networks. An important task in traffic analysis is the *shortest/fastest path computation*. Classical shortest path problem focuses on efficient computation in a large road network. However, historical traffic data may discover real fastest path and thus an adaptive fastest path computation method is proposed. Another interesting topic is to *predict destination* of moving vehicles. Methods based on Bayesian classification and frequent pattern have been developed to efficiently and effectively predict the recent or distant movement. In real life, people are also concerned with *road conditions*, such as hot/jammed roads and abnormal events on some road segments. To monitor road conditions, density-based routes clustering method is developed for hot routes discovery and temporal outlier detection in vehicle data is used to find abnormal road segments. Lastly, since road network naturally forms hierarchical structures and different granularity is embedded in temporal data, it is necessary to analyze moving object data in a *multidimensional way*, such as multidimensional traffic anomaly detection on highways and traffic cube and mining in traffic cube space.

In summary, this tutorial presents the state-of-the-art research on moving object data analysis including pattern mining, trajectory clustering, classification, outlier detection, and traffic analysis. It shows the confluence of multiple scientific and engineering disciplines, including data mining, database systems, geographic information system, statistical analysis, and machine learning, and links to multiple applications. We also discuss several promising research directions.