Chapter 2

A Query Language for Mobility Data Mining

Roberto Trasarti  
ISTI-CNR, Italy

Fosca Giannotti  
ISTI-CNR Italy, & Northeastern University, USA

Mirco Nanni  
ISTI-CNR, Italy

Dino Pedreschi  
University of Pisa, Italy, & Northeastern University, USA

Chiara Renso  
ISTI-CNR, Italy

ABSTRACT

The technologies of mobile communications and ubiquitous computing pervade society. Wireless networks sense the movement of people and vehicles, generating large volumes of mobility data, such as mobile phone call records and GPS tracks. This data can produce useful knowledge, supporting sustainable mobility and intelligent transportation systems, provided that a suitable knowledge discovery process is enacted for mining this mobility data. In this paper, the authors examine a formal framework, and the associated implementation, for a data mining query language for mobility data, created as a result of a European-wide research project called GeoPKDD (Geographic Privacy-Aware Knowledge Discovery and Delivery). The authors discuss how the system provides comprehensive support for the Mobility Knowledge Discovery process and illustrate its analytical power in unveiling the complexity of urban mobility in a large metropolitan area, based on a massive real life GPS dataset.

INTRODUCTION

Research on mobility data analysis has been recently fostered by the widespread diffusion of new techniques and systems for monitoring, collecting and storing location-aware data, generated by a wealth of wireless and mobile technologies, such as GPS positioning, mobile phones and sensor networks, tracking devices (Giannotti et al., 2008). These continuously feed massive repositories of spatio-temporally referenced data of moving objects, which call for suitable analytical methods for understanding mobile behavior. So far, research efforts have been largely aimed towards either
the definition of new movement model, or the development of solutions to algorithmic issues, to improve existing model-mining schemes in terms of effectiveness and/or efficiency. Unfortunately, discovering useful knowledge from these new forms of mobility data cannot be achieved by simply invoking an automated tool: as data miners know, successful analytics is the fruit of an overall knowledge discovery process, from raw data to knowledge. Figure 1 depicts the steps of the knowledge discovery process on movement data. Here, raw positioning data are collected from mobile devices and stored in the data repository. Trajectory data are then built, stored and analyzed by data mining algorithms to discover models hidden in the data. This process is typically iterative, since the composition of subsequent data mining methods is needed, both on data and model themselves, to obtain useful results. Finally, the extracted models have to be interpreted in order to be deployed by the final users.

The need of mastering the overall complexity of the knowledge discovery process led past research in the direction of inductive databases and data mining query languages (DMQL). Here, approaches provided various instances of querying and mining systems, all supporting the idea that discovering useful knowledge is a human-driven, and iterative and exploratory query process. Two main principles underlie this vision:

- **Persistence of data and models**: not only data, but also extracted models should be stored, in order to be further queried or mined (closure principle).
- **Expressiveness of the query language**: a high-level vision over data and models should be provided to the analyst.

The various DMQL proposals, described in the next section, refer to relational or transactional data and associated models; therefore, they are not directly exploitable for mobility knowledge discovery, given the very nature of movement of data and models. To bridge this gap, we designed and realized a comprehensive querying and mining

Figure 1. The Mobility Knowledge Discovery process
20 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the product's webpage: www.igi-global.com/chapter/query-language-mobility-data-mining/70791?camid=4v1


Related Content

A Method of Sanitizing Privacy-Sensitive Sequence Pattern Networks Mined From Trajectories Released

Referential Horizontal Partitioning Selection Problem in Data Warehouses: Hardness Study and Selection Algorithms

Efficient Algorithm for Mining High Utility Pattern Considering Length Constraints

A Data Mining-Based OLAP Aggregation of Complex Data: Application on XML Documents