Spatial Data Mining for Highlighting Hotspots in Personal Navigation Routes

Gabriella Schoier, University of Trieste, Italy
Giuseppe Borruso, University of Trieste, Italy

ABSTRACT

Rapid developments in the availability and access to spatially referenced information in a variety of areas have induced the need for better analytical techniques to understand the various phenomena. In particular, the authors’ analysis is an insight into a wealth of geographical data collected by individuals as activity dairy data. The attention is drawn on point datasets corresponding to GPS traces driven along a same route in different days. In this paper, the authors explore the presence of clusters along the route, trying to understand the origins and motivations behind that to better understand the road network structure in terms of ‘dense’ spaces along the network. Therefore, the attention is focused on methods to highlight such clusters and see their impact on the network structure. Spatial clustering algorithms are examined (DBSCAN) and a comparison with other non-parametric density based algorithm (Kernel Density Estimation) is performed. Different tests are performed over the urban area of Trieste (Italy), considering both multiple users and different origin/destination journeys.

Keywords: Activity Dairy Data, Clustering Algorithms, Density-Based Spatial Clustering of Applications with Noise (DBSCAN), GPS Traces, Kernel Density Estimation, Spatial Data Mining

INTRODUCTION

The rapid developments in the availability and access to spatially referenced information in a variety of areas have induced the need for better analytical techniques to understand the various phenomena. In particular spatial clustering algorithms, which groups similar spatial objects into classes, can be used for the identification of areas sharing common characteristics. The aim of this paper is to present a density based algorithm for the discovery of clusters of units in large spatial data sets. In particular the analysis represents a first insight into a wealth of geographical data collected by individuals as activity dairy data, these representing the routes a set of individuals drive in their daily movements for reasons connected to work, children picking, free time, rest. In this analysis the attention is drawn on point data sets corresponding to GPS traces driven along both a same route in different days and also considering a set of different paths recorded in different days and along different routes. Our aim here is to explore the presence

DOI: 10.4018/jdwm.2012070103
of clusters along the route, trying to understand the origins and motivations behind that in order to better understand the road network structure in terms of ‘dense’ spaces along the network; these representing road congestion, rather than the presence of junctions or other factors affecting individual mobility. In this paper the attention is therefore focused on methods to highlight such clusters and see their impact on the network structure.

Spatial clustering algorithms are examined (DBSCAN) and a comparison with other non-parametric density based algorithm (Kernel Density Estimation) is performed. Tests are performed on different journeys and by different users over the urban area of Trieste (Italy).

The paper is organized as follows.

In the paragraph “Spatial data mining” a survey of the characteristics of Data Mining techniques and the challenges of their application and performances in spatial environments are reported. A paragraph on “Time geography” follows, focusing on the definition of this research topic, in which people’s behaviors in space are also referred to a particular time and therefore on how today’s computational power, GIS software and location applications on personal devices to collect positional and temporal data allows the analysis of people’s movement over space and time, also rising new issues in terms of methods and instruments to be implemented. In the chapter on “Spatial clustering algorithms” the attention is drawn on the algorithms used in this paper for the analysis on individual personal data, in particular the DBSCAN algorithm and the Kernel Density Estimation. A gallery of applications of the two procedures follows in “Some comparisons between DBSCAN and KDE,” where the routes taken in different days are analyzed and compared in application of the two methods. In the “Conclusions” an overall comment is provided, while in the end of the paper some “Future research directions” are outlined.

**SPATIAL DATA MINING**

In recent years geographic data collection devices linked to location-aware technologies such as the global positioning system allow researchers to collect huge amounts of data. Other devices such as cell phones, in-vehicle navigation systems and wireless Internet clients can capture data on individual movement patterns. This explosive growth of spatial data and widespread use of spatial databases emphasize the need for the automated discovery of spatial knowledge.

The process of extracting information and knowledge from these massive geo-referenced databases is known as Geographic Knowledge Discovery (GKD) or Spatial Data Mining. It may be useful to understand spatial data, to discover relationships between spatial and non spatial data, to build knowledge-bases. This has a wide application in Geographic Information Systems (GIS), image analysis and other different areas where spatial data are used.

The nature of geographic entities, their complexity, relationships, and data means that standard Knowledge Discovery in Databases (KDD) or Data Mining techniques are not sufficient (Koperski, 1998) or at least their usefulness is limited. In fact the data inputs of Spatial Data Mining include extended objects such as points, lines, and polygons.

Specific reasons are the nature of geographic space, the complexity of spatial objects and relationships as well as their transformations over time, the heterogeneous and sometimes ill-structured nature of geo-referenced data, and the nature of geographic knowledge.

Spatial objects are embedded in a continuous space that serves as a measurement framework for all other attributes. This framework generates a wide spectrum of implicit distance, directional and topological relationships, particularly if the objects are greater than one dimension (such as lines, polygons...
Spatial OLAP and Map Generalization: Model and Algebra
Sandro Bimonte, Michela Bertolotto, Jérôme Gensel and Omar Boussaid (2012). International Journal of Data Warehousing and Mining (pp. 24-51).
www.igi-global.com/article/spatial-olap-map-generalization/61423?camid=4v1a

Spatial Clustering in SOLAP Systems to Enhance Map Visualization
Ricardo Silva, João Moura-Pires and Maribel Yasmina Santos (2012). International Journal of Data Warehousing and Mining (pp. 23-43).
www.igi-global.com/article/spatial-clustering-solap-systems-enhance/65572?camid=4v1a