Chapter 8
Geographic Knowledge Discovery in Multiple Spatial Databases

Tahar Mehenni
University Mohamed Boudiaf of M'sila, Algeria

ABSTRACT

Voluminous geographic data have been, and continue to be, collected from various Geographic Information Systems (GIS) applications such as Global Positioning Systems (GPS) and high-resolution remote sensing. For these applications, huge amount of data is maintained in multiple disparate databases and different in spatial data type, file formats, data schema, access mechanism, etc. Spatial data mining and knowledge discovery has emerged as an active research field that focuses on the development of theory, methodology, and practice for the extraction of useful information and knowledge from massive and complex spatial databases. This chapter highlights recent theoretical and applied research in geographic knowledge discovery and spatial data mining in a distributed environment where spatial data are dispersed in multiple sites. The author will present in this chapter, an overall picture of how spatial multi-database mining is achieved through several common spatial data-mining tasks, including spatial cluster analysis, spatial association rule and spatial classification.

INTRODUCTION

Geographic data consist of spatial objects and non-spatial description of these objects. Non-spatial description of spatial objects can be stored in a traditional relational database where one attribute is a pointer to spatial description of the object. Spatial data can be described using two different properties, geometric and topological. For example, geometric properties can be spatial location, direction area, perimeter, etc., whereas topological properties can be adjacency (object A is neighbor of object B), inclusion (object A is inside in object B), and others (Koperski, Adhikary, & Han, 1996).

A geographical database constitutes a spatio-temporal continuous recipient in which properties of a particular region, place or location are generally linked and explained in terms of the properties of

its neighborhood. Voluminous geographic data have been, and continue to be, collected by permanent efforts of scientific projects, government agencies, and private sectors. We now can obtain much more diverse, dynamic, and detailed data than ever possible before with modern data collection techniques, such as global positioning systems (GPS), high-resolution remote sensing, location-aware services and surveys, and internet-based volunteered geographic information (Goodchild & Yuan, 2007).

Facing the massive data that are increasingly available and the complex analysis questions that they may potentially answer, traditional spatial data analysis methods to find relationships and patterns are still having many limitations, because of their limited models and their inability to analyze newly emerged data types (such as surveillance videos and trajectories of moving objects) (Miller & Han, 2009).

Data Mining is the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data. Data mining methods (Fayyad, Piatetsky-Shapiro, & Smyth, 1996) are not suited to spatial data because they do not support location data or implicit relationships between objects.

Spatial data mining refers to the extraction of knowledge, spatial relationships, or other interesting patterns not explicitly stored in spatial databases. Hence, it is necessary to develop new methods including spatial relationships and spatial data handling. Calculating these spatial relationships is time consuming, and a huge volume of data is generated by encoding geometric location. Global performances will suffer from this complexity.

Knowledge discovery in spatial databases is the extraction of interesting spatial patterns and features, general relationships between spatial and nonspatial data, and other general data characteristics not explicitly stored in spatial databases. There is an urgent need for effective and efficient methods to extract unknown and unexpected information from datasets of unprecedentedly large size (e.g., millions of observations), high dimensionality (e.g., hundreds of variables), and complexity (e.g., heterogeneous data sources, space-time dynamics, multivariate connections, explicit and implicit spatial relations and interactions). To address these challenges, geographic knowledge discovery and spatial data mining has emerged as an active research field, focusing on the development of theory, methodology, and practice for the extraction of useful information and knowledge from massive and complex spatial databases (Koperski, Adhikary, & Han, 1996; Han, Koperski, & Stefanovic, 1997; Ester, Kriegel, & Sander, 2001; Chelghoum & Zeitouni, 2003; Shekhar, Zhang, Huang, & Vatsavai, 2003; Miller H. J., 2003; Miller & Han, 2009; (Leung, 2010; Perumal, Velumani, Sadhasivam, & Ramaswamy, 2015; Jassar & Dhindsa, 2015; Arvind, Jat, & Gupta; Zeitouni, 2002).

Spatial databases are traditionally meant for a particular field of application such as production of topographic maps, management of cadastral parcels, regional planning and so on. A same part of the real world can thus be described in several independent databases with different levels of detail and different points of views. Combining all the various aspects of each database by means of their integration is becoming a topical issue. However, spatial data distributed at multiple sites are difficult to aggregate into a single database for a variety of practical constraints, including dispersed data over many different geographic locations, security services and reasons of competition. Therefore, there is a great need for effective and efficient approaches to address these challenges.

Knowledge discovery in multiple spatial databases pose a research problem on data management and algorithms development. Due to the advances in communication technologies, data are distributed over the wired and wireless networks. How to discover spatial knowledge in a distributed computing environment is thus an area for further research. Algorithms for clustering, classification, associations

21 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/geographic-knowledge-discovery-in-multiple-spatial-databases/198550

Related Content

Contemporary Trends and Technologies in Research Libraries: An Overview
Neha Lataand Valentine Joseph Owan (2022). *Innovative Technologies for Enhancing Knowledge Access in Academic Libraries (pp. 40-56).*
www.irma-international.org/chapter/contemporary-trends-and-technologies-in-research-libraries/306427

Semantic Role Labeling Approach for Evaluation of Text Coherence
Mohamed H. Haggag (2013). *International Journal of Information Retrieval Research (pp. 59-77).*
www.irma-international.org/article/semantic-role-labeling-approach-for-evaluation-of-text-coherence/101360

Dynamic Data Retrieval Using Incremental Clustering and Indexing
Uma Priya Dand Santhi Thilagam P (2020). *International Journal of Information Retrieval Research (pp. 74-91).*
www.irma-international.org/article/dynamic-data-retrieval-using-incremental-clustering-and-indexing/257011

Role of Public Libraries in Bridging the Digital Divide
Owajeme Justice Ofuaand Ogochukwu Thaddaeus Emiri (2013). *Modern Library Technologies for Data Storage, Retrieval, and Use (pp. 185-193).*
www.irma-international.org/chapter/role-public-libraries-bridging-digital/73775

Ranking Documents Based on the Semantic Relations Using Analytical Hierarchy Process: Query Expansion and Ranking Process
Ali I. El-Dsouky, Hesham A. Aliand Rabab Samy Rashed (2018). *Information Retrieval and Management: Concepts, Methodologies, Tools, and Applications (pp. 1841-1859).*
www.irma-international.org/chapter/ranking-documents-based-on-the-semantic-relations-using-analytical-hierarchy-process/198628