Chapter 2.22
Mining Geo-Referenced Databases:
A Way to Improve Decision-Making

Maribel Yasmina Santos
University of Minho, Portugal

Luís Alfredo Amaral
University of Minho, Portugal

ABSTRACT

Knowledge discovery in databases is a process that aims at the discovery of associations within data sets. The analysis of geo-referenced data demands a particular approach in this process. This chapter presents a new approach to the process of knowledge discovery, in which qualitative geographic identifiers give the positional aspects of geographic data. Those identifiers are manipulated using qualitative reasoning principles, which allows for the inference of new spatial relations required for the data mining step of the knowledge discovery process. The efficacy and usefulness of the implemented system — PADRÃO — has been tested with a bank dataset. The results support that traditional knowledge discovery systems, developed for relational databases and not having semantic knowledge linked to spatial data, can be used in the process of knowledge discovery in geo-referenced databases, since some of this semantic knowledge and the principles of qualitative spatial reasoning are available as spatial domain knowledge.

INTRODUCTION

Knowledge discovery in databases is a process that aims at the discovery of associations within data sets. Data mining is the central step of this process. It corresponds to the application of algorithms for identifying patterns within data. Other steps are related to incorporating prior domain knowledge and interpretation of results.

The analysis of geo-referenced databases constitutes a special case that demands a particular approach within the knowledge discovery

Figure 1. Knowledge Discovery Process
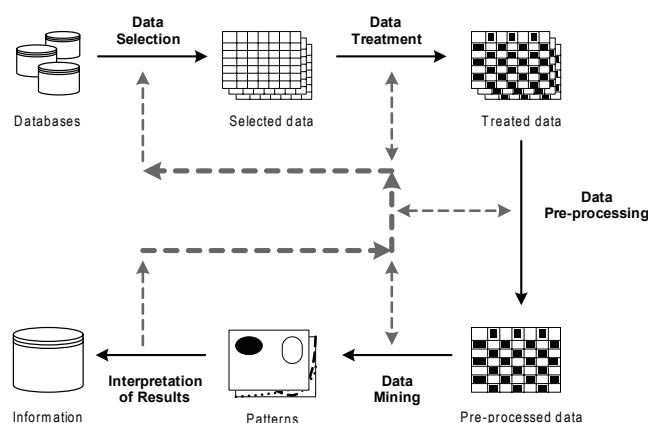
process. Geo-referenced data sets include allusion to geographical objects, locations or administrative sub-divisions of a region. The geographical location and extension of these objects define implicit relationships of spatial neighborhood. The data mining algorithms have to take this spatial neighborhood into account when looking for associations among data. They must evaluate if the geographic component has any influence in the patterns that can be identified.

Data mining algorithms available in traditional knowledge discovery tools, which have been developed for the analysis of relational databases, are not prepared for the analysis of this spatial component. This situation led to: (i) the development of new algorithms capable of dealing with spatial relationships; (ii) the adaptation of existing algorithms in order to enable them to deal with those spatial relationships; (iii) the integration of the capabilities for spatial analysis of spatial database management systems or geographical information systems with the tools normally used in the knowledge discovery process.

Most of the geographical attributes normally found in organizational databases (e.g., addresses) correspond to a type of spatial information, namely qualitative, which can be described using indirect positioning systems. In systems of spatial referencing using geographic identifiers, a position is referenced with respect to a real world location defined by a real world object. This object is termed a *location*, and its identifier is termed a *geographic identifier*. These geographic identifiers are very common in organizational databases, and they allow the integration of the spatial component associated with them in the process of knowledge discovery.

This chapter presents a new approach to the analysis of geo-referenced data. It is based on qualitative spatial reasoning strategies, which enable the integration of the spatial component in the knowledge discovery process. This approach, implemented in the PADRÃO system, allowed the analysis of geo-referenced databases and the identification of implicit relationships existing between the geo-spatial and non-spatial data.

The following sections, in outline, include: (i) an overview of the process of knowledge discovery and its several phases. The approaches usually followed in the analysis of geo-referenced databases are also presented; (ii) a description of qualitative spatial reasoning presenting its principles and the several spatial relations — direction, distance and topology. For the relations, an integrated spatial reasoning system was constructed and made available in the Spatial Knowledge Base of the PADRÃO

Related Content

Introduction of Item Constraints to Discover Characteristic Sequential Patterns

Shigeaki Sakurai (2019). *Emerging Perspectives in Big Data Warehousing (pp. 279-292).*

www.irma-international.org/chapter/introduction-of-item-constraints-to-discover-characteristic-sequential-patterns/231016

Robust Face Recognition for Data Mining

Brian C. Lovelland Shaokang Chen (2008). *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications (pp. 3621-3629).*

www.irma-international.org/chapter/robust-face-recognition-data-mining/7852

Mining Quantitative and Fuzzy Association Rules

Hong Shenand Susumu Horiguchi (2005). *Encyclopedia of Data Warehousing and Mining (pp. 815-819).*

www.irma-international.org/chapter/mining-quantitative-fuzzy-association-rules/10709

Data Mining and Homeland Security

Jeffrey W. Seifert (2008). *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications (pp. 3630-3638).*

www.irma-international.org/chapter/data-mining-homeland-security/7853

A Haplotype Analysis System for Genes Discovery of Common Diseases

Takashi Kido (2008). *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications (pp. 1674-1687).*

www.irma-international.org/chapter/haplotype-analysis-system-genes-discovery/7723