

Chapter IX

Enhancing the Process of Knowledge Discovery in Geographic Databases Using Geo-Ontologies

Vania Bogorny

Universidade Federal do Rio Grande do Sul (UFRGS), Brazil, and Transnational University of Limburg, Belgium

Paulo Martins Engel

Universidade Federal do Rio Grande do Sul (UFRGS), Brazil

Luis Otavio Alavares

Universidade Federal do Rio Grande do Sul (UFRGS), Brazil

ABSTRACT

This chapter introduces the problem of mining frequent geographic patterns and spatial association rules from geographic databases. In the geographic domain most discovered patterns are trivial, non-novel, and noninteresting, which simply represent natural geographic associations intrinsic to geographic data. A large amount of natural geographic associations are explicitly represented in geographic database schemas and geo-ontologies, which have not been used so far in frequent geographic pattern mining. Therefore, this chapter presents a novel approach to extract patterns from geographic databases using geo-ontologies as prior knowledge. The main goal of this chapter is to show how the large amount of knowledge represented in geo-ontologies can be used to avoid the extraction of patterns that are previously known as noninteresting.

INTRODUCTION

Knowledge discovery in databases (KDD) is the nontrivial process of identifying valid, novel, potentially useful and ultimately understandable patterns from data (Fayyad et al., 1996). In frequent pattern mining (FPM), which is the essential role in mining associations, one of the main problems is the large amount of generated patterns and rules. In geographic databases this problem increases significantly because most discovered patterns include well-known natural associations intrinsic to geographic data. While in transactional databases items are supposed to be independent from each other (e.g., milk, cereal, bread), independently of their meaning, in geographic databases a large amount of data are semantically dependent (e.g., island *within* water).

Geographic dependences are semantic constraints that must hold in geographic databases (GDB) to warrant the consistency of the data (e.g., island must be completely located inside a water body). They are part of the concept of geographic data and are explicitly represented in geo-ontologies. Without considering semantics of geographic data, the same geographic dependences explicitly represented in geo-ontologies and geographic database schemas are unnecessarily extracted by association rule mining algorithms and presented to the user.

Geographic dependences produce two main problems in the process of mining spatial association rules:

- a. **Data preprocessing:** A large computational time is required to preprocess GDB to extract spatial relationships (e.g., *intersection* between districts and water bodies). The spatial join (Cartesian product) operation, required to extract spatial relationships, is the most expensive operation in databases and the processing bottleneck of spatial data analysis and knowledge discovery.
- b. **Frequent pattern and association rule generation:** A large number of patterns and spatial association rules without novel, useful, and interesting knowledge is generated (e.g., *is_a(Island) → within (Water)*).

Aiming to improve geographic data preprocessing and eliminate well-known geographic dependences in geographic FPM in order to generate more interesting spatial association rules (SAR), this chapter presents a unified framework for FPM considering the semantics of geographic data, using geo-ontologies. While dozens of spatial and nonspatial FPM algorithms define syntactic constraints and different thresholds to reduce the number of patterns and association rules, we consider *semantic knowledge constraints* (Bogorny et al., 2005b), and eliminate the exact sets of geographic objects that produce well-known patterns (Bogorny et al., 2006b, 2006c).

The main objective of this chapter is to show the important role that ontologies can play in the knowledge discovery process using the FPM technique. The focus addresses the use of semantic knowledge stored in ontologies to reduce uninteresting patterns, but not to create ontologies for data mining.

The remainder of the chapter is organized as follows: Section 2 presents some background concepts about geographic data, spatial relationships, spatial integrity constraints, and geo-ontologies. Section 3 introduces the concepts of frequent patterns and spatial association rules, the problem generated by geographic dependences in both data preprocessing and spatial association rule mining, and what has been done so far to alleviate this problem. Section 4 presents a framework to improve geographic data preprocessing and spatial association rule mining using geo-ontologies. Experiments are presented to show the significant reduction in the number of frequent patterns and association rules. Section 5 presents future trends and Section 6 concludes the chapter.

20 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/enhancing-process-knowledge-discovery-geographic/7577

Related Content

A Dynamic Privacy Manager for Compliance in Pervasive Computing

Riccardo Bonazzi, Zhan Liu, Simon Ganière and Yves Pigneur (2013). *Data Mining: Concepts, Methodologies, Tools, and Applications* (pp. 793-815).

www.irma-international.org/chapter/dynamic-privacy-manager-compliance-pervasive/73471

Multi-Criteria Decision Making in Manufacturing Systems: Identification of Critical Factors for Establishing a Smart Factory Using ISM and MICMAC Approach

Hande Erdoan Aktan and Ömür Tosun (2019). *Optimizing Big Data Management and Industrial Systems With Intelligent Techniques* (pp. 80-107).

www.irma-international.org/chapter/multi-criteria-decision-making-in-manufacturing-systems/218741

Boat Detection in Marina Using Time-Delay Analysis and Deep Learning

Romane Scherrer, Erwan Aulnette, Thomas Quiniou, Joël Kasarherou, Pierre Kolband Nazha Selmaoui-Folcher (2022). *International Journal of Data Warehousing and Mining* (pp. 1-16).

www.irma-international.org/article/boat-detection-in-marina-using-time-delay-analysis-and-deep-learning/298006

TripRec: An Efficient Approach for Trip Planning with Time Constraints

Heli Sun, Jianbin Huang, Xinwei She, Zhou Yang, Jiao Liu, Jianhua Zou, Qinbao Song and Dong Wang (2015). *International Journal of Data Warehousing and Mining* (pp. 45-65).

www.irma-international.org/article/triprec/122515

A New Approach for Supervised Dimensionality Reduction

Yinglei Song, Yongzhong Li and Junfeng Qu (2018). *International Journal of Data Warehousing and Mining* (pp. 20-37).

www.irma-international.org/article/a-new-approach-for-supervised-dimensionality-reduction/215004