

Chapter 58

SPAM:

An Effective and Efficient Spatial Algorithm for Mining Grid Data

Ritu Chauhan
Amity University, India

Harleen Kaur
Hamdard University, India

ABSTRACT

The tremendous increase in spatial database technology has created furious interest among the researchers to adopt new methodologies for discovery of interesting patterns among large databases. But the raw data gathered from various resources such as Geographic Information Systems (GIS), business organizations, medical databases, climatic, market survey, remote sensing and several other resources might consist of data, which can be relevant, irrelevant or noisy in nature. However, retrieval of patterns from such databases can lead to serious concerns, which can frame inconsistent or irrelevant futuristic results. To deal with such issues, feature selection techniques are adopted to remove irrelevant, redundant and noisy features. Our approach focuses on retrieval of effective and efficient spatial clusters from large number of medical databases. In this chapter, we have defined our novel framework SpaGrid and SPAM algorithm to retrieve clusters of variant shape and size from large databases. The application of our framework is used with spatial medical databases where the implementation details are discussed with Matlab 7.1.

INTRODUCTION

In past decades, there has been explosive growth in spatial database technology generated from various sources. Such databases are usually evolved from numerous domains such as medical technology, business oriented organizations, market based analysis, social analysis, science exploration, geographical information studies, and several other automation technologies for retrieval and storage of data. There are constant efforts laid down by researchers and scientists throughout the globe to retrieve hidden information from large spatial databases. However, spatial data analysis tends to be major necessity for discovery of undefined rules and patterns from large spatial databases.

DOI: 10.4018/978-1-5225-5191-1.ch058

In pervasive term Spatial databases comprise of spatial as well as non-spatial attributes. Spatial attributes consist of features related to specific areas or regions, whereas non-spatial attributes are height, weight, temperature, etc. of such areas. There exist general relationship characteristics between spatial and non-spatial attributes for extraction of interesting spatial patterns from raw data (Lui & Han, 1993). Spatial data consists of data related to medical images which include thematic maps of specific area, remote sensing and several other sources where spatial data is bonded with non-spatial attributes of specific location (Lui & Han, 1993; Rigaux & Scholl, 2002; Zaiane, 2002). However, the discrete nature of spatial data can be represented in different form such as lattice data which consists of regular or irregular data units, the second representation of spatial data includes point pattern data which occurs in space at different location and finally geostatistical data which is associated with continuous variation of data in space (Cressie, 1991).

There exists different representation of spatial data in space such as raster and vector representation. The Raster data representation divides the space into rectangular cells and each cell corresponds to some value such as maximum, minimum or entire value of cell is made into consideration. They are represented as points with respect to collection of area in specific space, whereas vector data is represented by line, point and polygon of specific area. Point belonging to these spatial objects represents spatial data in space. The comparison between raster and vector data is made, raster data is applied with simpler data structure as spatial objects are divided into smaller subsets. Vector data can be easily stored, transformed and rotated hence can efficiently store without the high memory requirement. Therefore, it was found that vector data tends to work more efficiently than raster data.

Such databases are multidimensional in nature; hence proper spatial data structures are required for storage and retrieval of raw data. Therefore, discovery of knowledge should be formed as an automated process for retrieval of information from spatial databases (Seeger & Kriegel, 1988).

Spatial data mining is the process to discover hidden and unknown patterns from spatial databases. In particular, it has been defined as the process for discovering interesting relationships and characteristics among the spatial databases (Ng & Han, 1994; Laurini & Thompson, 1992; Fayyad et al., 1996; Kaur et al., 2010). The urge for spatial data mining exists due to extensive large numbers of spatial databases collected everyday through remote sensing, NASA satellite, climate conditions, which contain hidden relevant patterns which cannot be synthesized through human capabilities all alone.

In our proposed approach, we have developed an integrated framework SpaGrid for retrieval of spatial clusters from large spatial databases. The focus is to generate a spatial environment to handle complex and large databases. The framework involves an integrated approach, which is compromised of feature selection techniques; storage and retrieval of data through a specific spatial data structure, then propose a spatial clustering algorithm for retrieval of knowledgeable spatial clusters.

The developed framework is highly flexible and involves several approaches, which can address complex issues to demonstrate and discover spatial clusters for knowledge discovery process. Our chapter focuses on three-dimensional clustering of spatial objects to discover clusters of variant shape and size. To achieve our current objective, Spatial Algorithm with Merging (SPAM) has been proposed to determine spatial clusters of variant shape and size. The visualization and implementation of these algorithms are performed in Matlab 7.1 environment. However, contribution of our work focuses on the field of spatial data mining, especially spatial clustering on medical databases, i.e. the clusters are automatically generated with the help of proposed clustering algorithms.

13 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/spam/198601

Related Content

Bio-Inspired Algorithms for Ecosystem Data Analysis

Mohamed Elhadi Rahmani (2018). *Handbook of Research on Biomimicry in Information Retrieval and Knowledge Management* (pp. 225-250).

www.irma-international.org/chapter/bio-inspired-algorithms-for-ecosystem-data-analysis/197704

Mechanism for Crawling, Filtering, and Presenting Opinionated Content on Online Products to the Customers

Rosy Madaan (2023). *Applying AI-Based IoT Systems to Simulation-Based Information Retrieval* (pp. 169-182).

www.irma-international.org/chapter/mechanism-for-crawling-filtering-and-presenting-opinionated-content-on-online-products-to-the-customers/322856

Nakhasys: An ML-Based Disease Diagnosing Application

Poonam Tanwar (2023). *Applying AI-Based IoT Systems to Simulation-Based Information Retrieval* (pp. 90-103).

www.irma-international.org/chapter/nakhasys/322852

Ensemble of Support Vector Machine and Ontological Structures to Generate Abstractive Text Summarization

Amita Arora (2022). *International Journal of Information Retrieval Research* (pp. 1-24).

www.irma-international.org/article/ensemble-of-support-vector-machine-and-ontological-structures-to-generate-abstractive-text-summarization/300294

Video Presentation Model

Hun-Hui Hsu, Yi-Chun Liao, Yi-Jen Liu and Timothy K. Shih (2005). *Video Data Management and Information Retrieval* (pp. 177-192).

www.irma-international.org/chapter/video-presentation-model/30765