SPATIAL DATA MINING - TECHNIQUES TRENDS AND ITS APPLICATIONS

N.Sumathi
Senior Lecturer
K.S.Rangasamy College of Technology
Tiruchengode-637 215

R.Geetha
Senior Lecturer
K.S.Rangasamy College of Technology
Tiruchengode-637 215

Dr. S. Sathiya Bama
Professor
K.S.Rangasamy College of Technology
Tiruchengode-637 215

Abstract

Public and private organizations have legacy or operational spatial databases or non-spatial databases, which are also somehow linked to a spatial database or a spatial meaning. In addition to mission related databases, these organizations either have or access several databases comprising such as census, economic, security, image, multimedia, statistical information for planning, intelligence, decision and policy making. The extended size of these data sets makes it difficult to search for meaningful patterns or relationships among data.

Spatial data mining is the process of discovering interesting and previously unknown, but potentially useful patterns from large spatial datasets. This paper focuses on Techniques and the unique features that distinguish spatial data mining from classical data mining, Finally it identify areas of spatial data mining where further research is needed.

Keywords: Spatial data mining, Clustering and Outlier Detection, Association and Co-Location, Classification, Trend-Detection, Research needs.

1. Introduction

The explosive growth of spatial data and widespread use of spatial databases emphasize the need for the automated discovery of spatial knowledge. Spatial data mining is the process of discovering interesting and previously unknown, but potentially useful patterns from spatial databases. The complexity of spatial data and intrinsic spatial relationships limits the usefulness of conventional data mining techniques for extracting spatial patterns. Efficient tools for extracting information from geo-spatial data are crucial to organizations which make decisions based on large spatial datasets, including NASA, the National Imagery and Mapping Agency (NIMA), the National Cancer Institute (NCI), and the United States Department of Transportation (USDOT). These organizations are spread across many application domains including ecology and environmental management, public safety, transportation, Earth science, epidemiology, and climatology.

1.1 Spatial Data Mining Tasks

Basic tasks of spatial data mining are:

1.1.1 Classification – finds a set of rules which determine the class of the classified object according to its attributes

1.1.2 Association rules – find (spatially related) rules from the database. Association rules describe patterns, which are often in the database.

1.1.3 Characteristic rules – describe some part of database e.g. ”bridge is an object in the place where a road crosses a river.”

1.1.4 Discriminate rules – Describe differences between two parts of database e. g. find differences between cities with high and low unemployment rate.

1.1.5 Clustering – Groups the object from database into clusters in such a way that object in one cluster are similar and objects from different clusters are dissimilar.

1.1.6 Trend detection – Finds trends in database. A trend is a temporal pattern in some time series data. A spatial trend is defined as a pattern of change of a non-spatial attribute in the neighborhood of a spatial object.

2. Spatial Data Mining Techniques

There is no unique way of classifying SDM techniques. Various kinds of patterns can be discovered from databases and can be presented in different forms. Based on general data mining, tasks can be classified into two main categories: descriptive data mining and predictive data mining.

The former concisely describes the behavior of datasets and presents interesting general properties of the data. Whereas the latter attempts to construct models that tend to help predicting the behavior of the new datasets. This paper focuses on the organization of the particular spatial data mining techniques as

- **. Clustering and Outlier Detection**
- **. Association and Co-Location**
- **. Classification**
- **. Trend-Detection**

2.1 Clustering And Outlier Detection

Spatial clustering is a process of grouping a set of spatial objects into groups called clusters. Objects within a cluster show a high degree of similarity, whereas the clusters are as much dissimilar as possible.

Clustering is a very well known technique in statistics and clustering algorithm to deal with the large geographical datasets. Clustering algorithms can be separated into four general categories: partitioning method, hierarchical method, density-based method and grid-based method. The categorization is based on different cluster definition techniques.

2.1.1 Partitioning Method

Partitioning method partitioning algorithm organizes the objects into clusters such that the total deviation of each object from its cluster center is minimized. At the beginning each object is classified as a single cluster. In the next steps, all data points are iteratively reallocated to every cluster until a stopping criterion is met. K-Means is commonly used fundamental partitioning algorithm.

2.1.2 Hierarchical Method

Hierarchical method hierarchically decomposes the dataset by splitting or merging all clusters until a stopping criterion is met. Some of the recently used hierarchical clustering algorithms are Balanced Iterative Reducing and Clustering using Hierarchies and Clustering Using Representatives.

2.1.3 Density-Based Method

The method regards clusters as dense regions of objects that are separated by regions of low density (representing noise). In contrast to partitioning methods, clusters of arbitrary shapes can be discovered. Density-based methods can be used to filter out noise and outliers.

2.1.4 Grid-Based Method

Grid-based clustering algorithms first quantize the clustering space into a finite number of cells and then perform the required operations on the grid structure. Cells that contain more than a certain number of points are treated as dense. The main advantage of the approach is its fast processing time, since the time is independent on the number of data objects, but dependent on the number of cells.

2.2 Association and Co-Location

When performing clustering methods on the data, we can find only characteristic rules, describing spatial objects according to their non-spatial attributes. In many situations we want to discover spatial rules that associate one or more spatial objects with others.

However, one of the biggest research challenges in mining association rules is the development of methods for selecting potentially interesting rules from among the mass of all discovered rules.

2.3 Classification

Every data object stored in a database is characterized by its attributes. Classification is a technique, which aim is to find rules that describe the partition of the database into an explicitly given set of classes. Classification is considered as predictive spatial data mining, because we first create a model according to which the whole dataset is analyzed.

2.4 Trend Detection

A spatial trend is a regular change of one or more non-spatial attributes when spatially moving away from a start object. Therefore, spatial trend detection is a technique for finding patterns of the attribute changes with respect to the neighborhood of some spatial object.

3. Spatial Data Mining Trends

3.1 Data Input

The data inputs of spatial data mining have two distinct types of attributes: non-spatial attribute and spatial attribute. Non-spatial attributes are used to characterize non-spatial features of objects, such as name, population, and unemployment rate for a city. Spatial attributes are used to define the spatial location and extent of spatial objects. The spatial attributes of a spatial object most often include information related to spatial locations, e.g., longitude, latitude and elevation, as well as shape.

3.2 Computational Process

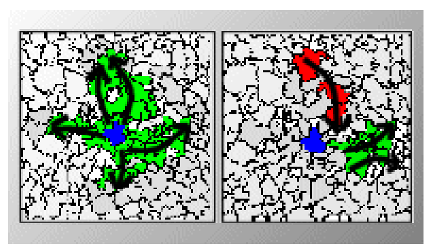
In spatial data mining, spatial autocorrelation and low dimensionality in space provide more opportunities to improve computational efficiency than classical data mining. NASA Earth observation systems currently generate a large sequence of global snapshots of the Earth, including various atmospheric, land, and ocean measurements such as sea surface temperature, pressure, precipitation, and net primary production.

4. Spatial Data Mining Applications

4.1 Spatial Trend Detections in GIS

Spatial trends describe a regular change of non-spatial attributes when moving away from certain start objects. Global and local trends can be distinguished. To detect and explain such spatial trends, e.g. with respect to the economic power, is an important issue in economic geography.

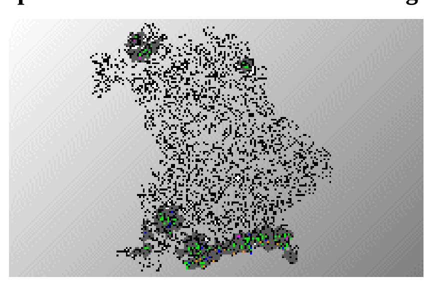
Fig.1 Spatial Trend Detections in GIS



4.2 Spatial Characterization of Interesting Regions

Another important task of economic geography is to characterize certain target regions such as areas with a high percentage of retirees. Spatial characterization does not only consider the attributes of the target regions but also neighboring regions and their properties.

Fig.2 Spatial Characterization of Interesting Regions



5. Research Needs

In this section, we discuss some areas where further research is needed in spatial data mining.

- **Comparison of classical data mining techniques with spatial data mining techniques**
  Existing literature does not provide guidance regarding the choice between classical data mining techniques and spatial data mining techniques to mine spatial data. Therefore new research is needed to compare the two sets of approaches in effectiveness and computational efficiency.
- **Modeling semantically rich spatial properties, such as topology**
  Spatial connectivity and other complex spatial topological relationships in spatial networks are difficult to model using the continuity matrix. Research is needed to evaluate the value of enriching the continuity matrix beyond the neighborhood relationship
- **Improving computational efficiency**
  Mining spatial patterns is often computationally expensive. For example, the estimation of the parameters for the spatial autoregressive model is an order of magnitude more expensive than that for the linear regression in classical data mining. Research is needed to reduce the computational costs of spatial data mining algorithms by a variety of approaches including the classical data mining algorithms as potential filters or components.
- **Preprocessing spatial data**
  There is a need for preprocessing techniques for spatial data to deal with problems such as treatment of missing location information and imprecise location specifications, cleaning of spatial data, feature selection and data transformation.

5. Conclusion

This paper presented the techniques of spatial data mining in the following four categories Clustering and Outlier Detection, Association and Co-Location, Classification and Trend-Detection. It also discussed some trends and applications of spatial data mining. Finally, it identified research needs for spatial data mining.

References

[1] Anselin, L. 1994. Exploratory Spatial Data Analysis and Geographic Information Systems. In Painho, M., ed., New Tools for SpatialAnalysis, 45-54.

[2] Anselin, L. 1995. Local Indicators of Spatial Association: LISA.Geographical Analysis 27(2):93-115.

[3] Barnett, V., and Lewis, T. 1994. Outliers in Statistical Data. John Wiley, 3 rd edition.

[4] Besag, J. 1974. Spatial Interaction and Statistical Analysis of Lattice Systems. Journal of Royal Statistical Society: Series B 36:192-36.