International Journal of Engineering Science Invention (IJESI)
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org || PP. 42-45

A Survey Report On Spatial Data Mining

G.PARTHIBAN
[1](Department of Computer Applications, AAGAC ,villupuram, India)

Abstract: *Spatial data mining and knowledge discovery (SDMKD) is the efficient extraction of hidden, implicit, interesting, previously unknown, potentially useful, ultimately understandable, spatial or non-spatial knowledge (rules, regularities, patterns, constraints) from incomplete, noisy, fuzzy, random and practical data in large spatial databases. It is a confluence of databases technology, artificial intelligence, machine learning, probabilistic statistics, visualization The algorithm are classified into four category, i) Clustering and outlier detection, ii)Association and co-location method, iii)Trend detection and, iv) Classification.This paper focus only on the survey report of spatial data mining.*
Keywords: *Clustering, knowledge discovery, Outliers detection, Non-spatial knowledge, Spatial*

I. Introduction

It is a process of discovering interesting and previously unknown but potentially useful pattern from large spatial data set. It refers to the extraction of knowledge, spatial relationship or other interesting pattern not explicitly stored in spatial database. Extracting interesting and useful pattern from spatial data set. Storing of the data input in spatial dataset is not analogous to traditional data set. It is more difficulty than extracting corresponding pattern from traditional numeric and categorical data due to complexity of spatial data type, spatial relationship and spatial auto correction.Efficient tools for extracting information from geo-spatial data are NASA, the National Imagery and Mapping Agency (NIMA), National Cancer Institute (NCI), and the United States Department of Transport (USDOT). A spatial database is the database which has been specially optimized to store data pertaining to objects in the real world. In other words spatial data is the data which represents objects in geometric space. The objects are stored in database in the form of lines, points and polygons. A Relational Database Management System (RDBMS) with additional features can support spatial databases which are extensively used in environmental studies, Global Positioning System (GPS), and Geographic Information System (GIS). Spatial Data Mining (SDM) is a process of discovering trends or patterns from large spatial databases that hold geographical data. Objects in space such as roads, rivers, forests, deserts, buildings, cities etc., are stored in spatial database. Spatial databases are so complex and make the SDM more difficult when compared with traditional databases. The major applications of SDM are related to co-location mining, spatial outlier detection and location prediction. PostGIS, Microsoft SQL Server, Oracle Spatial, Spatiality etc., are the products available for building spatial databases. This paper overviews the applications and uses of spatial mining in health.10)

II. Essence Of Approach

2.1 Basic Facts

Data mining has attracted a great attention in the information industry and in society as a whole in recent years, due to wide availability of huge amount of data and the imminent need for turning such data into useful information and knowledge. The information and knowledge gained can be used for application ranging from market analysis, fraud detection, to production control, disaster management and science exploration. Data mining can be viewed as a result of the natural evolution of information technology. The database system industry has witnessed an evolutionary path in the development of various functionalities: data collection and database creation, database management (including data storage and retrieval, and database transaction processing and advance data analysis Knowledge discovery as a process consists of an iterative sequence of following steps:

- *Data cleaning,* that is, to remove noise andinconsistent data.
- *Data integration,* that is, where multiple data sourcesare combined.
- *Data selection,* that is, where data relevant to theanalysis task are retrieved from the database.
- *Data transformation,* that is, where data aretransformed or consolidated into forms appropriate for mining by performing summary or aggregation operations.
- *Data mining,* that is, an essential process whereintelligent methods are applied in order to extract the data patterns.

- *Knowledge presentation*, that is, where visualizationand knowledge representation techniques are used to present the mined knowledge to the user.

2.2Spatial Data Mining

Spatial data and previously unknown, but potentially mining is the process of discovering interesting useful patterns from large spatial datasets. Extracting interesting and useful patterns from spatial datasets is more difficult than extracting the corresponding patterns from traditional numeric and categorical data due to the complexity of spatial data types, spatial relationships, and spatial autocorrelation. 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 of 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.

2.2.1Characteristics of Spatial Data Mining

- It is mainly used for auto correlation
- Patterns usually have to be defined in the spatial attribute subspace and not in the complete attribute space.
- Patterns usually have to be defined in the spatial attribute subspace and not in the complete attribute space.

2.2.2 Applications Of Spatial Data Mining

- identifying Fire Hot Spot in forest
- Railway Geographic Information System
- Evaluation of forest extent changes
- Analyzing distribution of regional economy
- Grading of Agriculture through Spatial Data Mining
- Agriculture Crop Yield Prediction

2.2.3Application In Visual Data Mining

For data mining of large data sets to be effective, it is also important to include humans in the data exploration process and combine their flexibility, creativity, and general knowledge with the enormous storage capacity and computational power of today's computers. Visual data mining applies human visual perception to the exploration of large data sets. Presenting data in an interactive, graphical form often fosters new insights, encouraging the formation and validation of new hypotheses to

2.2.4 Spatial Classification

The task of classification is to assign an object to a class from a given set of classes based on the attribute values of this object. In spatial classification the attribute values of neighboring objects are also considered. The task of classification is to assign an object to a class from a given set of classes based on the attribute values of this object. In spatial classification the attribute values of neighboring objects are also considered. The end of better problem solving and gaining deeper domain knowledge. Visual data mining often follows a three step process: Overview first, zoom and filter, and then details-on demand. Some of the key advantages of visual data exploration over automatic data mining techniques alone are yields results more quickly, with a higher degree of user satisfaction and confidence in findings are vague, because the analyst guides the search and can

- shift or adjust goals on the fly
- can deal with highly non-homogeneous and noisy data
- can be intuitive and require less understanding of complex
- Mathematical or statistical algorithms or parameters.
- an provide a qualitative overview of the data, allowing unexpected phenomena.

2.2.5 Spatial Neighborhood Relations

There are three basic types of spatial relations: topological, distance and direction relations which may be combined by logical operators to express a more complex neighborhood relation. Spatial objects such as points, lines, polygons or polyhedrons are all represented by a set of points. For example, a polygon can be represented by its edges (vector representation) or by the points contained in its interior, e.g. the pixels of an object in a raster image (raster representation). Topological relations are based on the boundaries, interiors and complements of the two related objects and are invariant under transformations which are continuous, one-one, onto and whose inverse is continuous.

III. Figures and Tables

3.1 Architecture of Spatial Data Mining

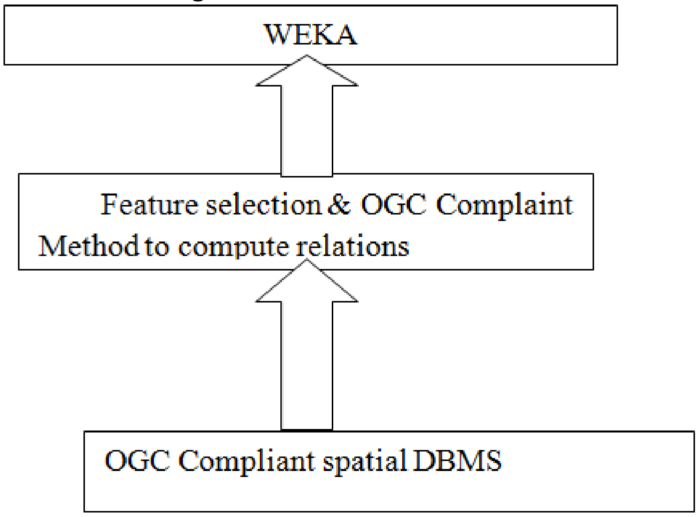

Fig .1 Architecture of Spatial Data Mining

In this architecture shows the data belonging to preprocess spatial data to materialize spatial features .This method is used to compute spatial relationship to create flat file of data.

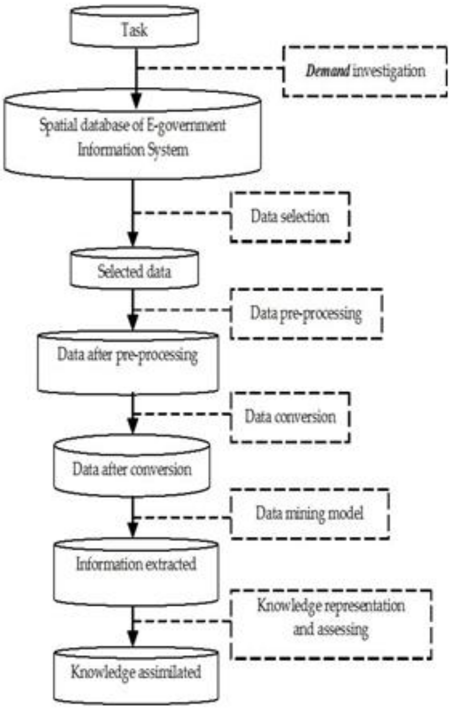

Fig 1.1 Task of Spatial Mining

3.2 Techniques Used in SDMKD

SDMKD is an interdisciplinary subject, there are various techniques associated with the abovementioned different knowledge (Li et al., 2002). They may include, probability theory, evidence theory (Dempster-Shafer), spatial statistics, fuzzy sets, cloud model, rough sets, neural network, genetic algorithms, decision tree, exploratory learning, inductive learning, visualization, spatial online analytical mining (SOLAM)

3.2 Summary of Results

Agriculture: It is used to estimate crop yielding and damages caused by pesticide

Geo Miner: A prototype of a spatial data mining system. The system design includes a graphical user interface (GUI) component for data visualization, modules for performing exploratory data analysis (EDA) and spatial data mining, and a spatial database server.
Weather Condition: It is used to find weather forecasting by using frequent pattern growth algorithm.
Railway System: An IRGIS uses intelligent spatial data mining to discover the association rules hidden in the vast amount of railway data

3.3 Application domains of Spatial Data Mining

It is used to Scale up secondary spatial (statistical) analysis to very large datasets. Describe/explainslocations of human settlements in last 5000 years. Find cancer clusters to locate hazardous environments. Prepare land-use maps from satellite imagery. Predict habitat suitable for endangered species. Find groups of co-located geographic features

IV. Conclusion

This paper overviews the different applications used in spatial data mining and its uses in medical industries in detail. Finally, it identified New techniques are needed for SDM due to Spatial Auto-correlation Importance of non-point data types (e.g. polygons)

References

[1]. N.Sumathi, R.Geetha and Dr.S.SathyaBama "Spatial Data Mining Techniques,Trends& its Applications", International Journal Of Computer Applications" (0975-887),Vol-1,No.4,28-30,Oct-Dec 2008
[2]. D.Rajesh "Application of Spatial Data Mining for Agriculture" International Journal Of Computer Applications (0975-887), Vol-15, No.2,Feb-2011
[3]. AakumuriManjula, Dr.G.Narsimha "A Review of Spatial Data Mining Methods and Applications" International Journal of Computer Engineering and Applications, Volume VII, Issue I, PartII, July 14.
[4]. AmruthaA.Taksanda, P.S.Mohod "Application of WeatherFore casting Using Frequent Pattern Growth Algorithm "Internal Journal of Science and Research.