International Journal of Engineering Research and Development
e-ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com
Volume 10, Issue 4 (April 2014), PP.47-52

Soft Spatial Query Processing in Spatial Databases-A Case Study

J RAJANIKANTH[1], Dr. T.V. RAJINIKANTH[2]
[1]*Assistant Professor, Dept of CSE, S.R.K.R Engg. College, Bhimavaram, A.P. INDIA.*
[2]*Professor & HOD, Department of IT, GRIET, Hyderbad, AP, INDIA.*

Abstract:- Knowledge discovery in databases (KDD) is an important task in spatial databases since both, the number and the size of such databases are rapidly growing. This paper introduces a set of basic operations which should be supported by a spatial database system (SDBS). Here, it will go on to describe the research that is currently going on in this area, pointing out the different areas, tasks etc., Finally this will helps the researchers to develop better techniques in spatial databases.

Keywords:- KDD, Database Model, SDBS.

I. INTRODUCTION

Spatial Database Systems are database systems for the management of spatial data. Both, the number and the size of spatial databases are rapidly growing in applications such as geomarketing, traffic control and environmental studies. This growth by far exceeds human capacities to analyze the databases in order to find implicit regularities, rules or clusters hidden in the data.

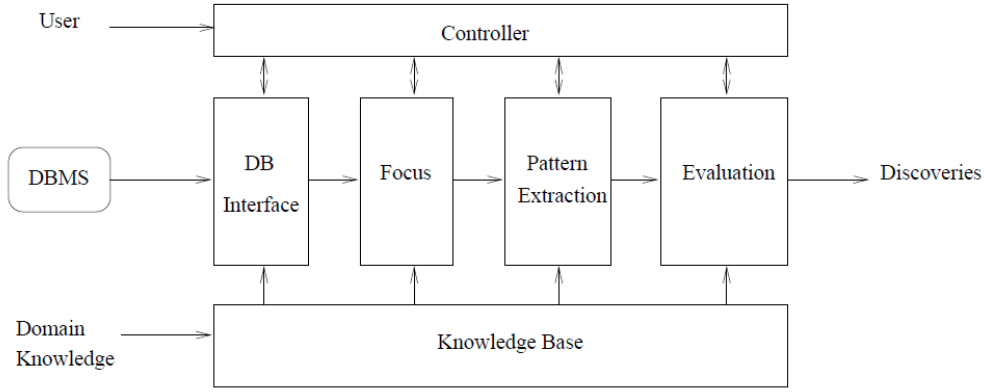

Fig 1 KDD architecture

Therefore, automated knowledge discovery becomes more and more important in spatial databases. The generalization based knowledge discovery requires the existence of background in the form of concept hierarchies. In the case of spatial databases, there can be two kinds of concept hierarchies, no-spatial and spatial. Concept hierarchies can be explicitly given by the experts, or in some cases they can generate automatically by data analysis [1-5]. Architecture for KDD system is shown in Fig 1

1 Spatial Data Mining Structure

The spatial data mining can be used to understand spatial data, discover the relation between space and the non space data, set up the spatial knowledge base, excel the query, reorganize spatial database and obtain concise total characteristic etc.. The system structure of the spatial data mining can be divided into three layer structures mostly, as shown in Figure 2[6-9].

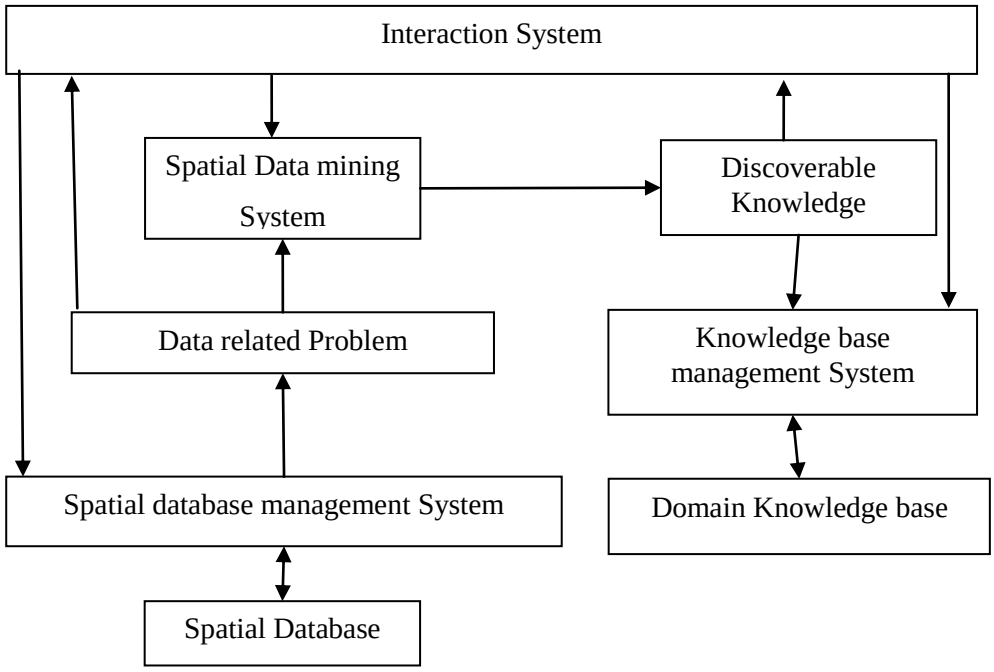

Fig 2 structure of spatial mining

The customer interface layer is mainly used for input and output, the miner layer is mainly used to manage data, select algorithm and storage the mined knowledge, the data source layer, which mainly includes the spatial database and other related data and knowledge bases, is original data of the spatial data mining.

II. PRIMITIVES OF SPATIAL DATA MINING

Rules: There are several kinds of rules can be discovered from databases in general. For example characteristic rules, discriminate rules, association rules, or deviation and evaluation rules can be mined [6-9]. A Spatial characteristic rule is a general description of the spatial data. For example, a rule describing the general price range of houses in various geographic regions in a city is a spatial characteristic rule. A discriminate rule is general description of the features discriminating or contrasting a class of spatial data from other classes like the comparison of price ranges of houses in different geographical regions. A spatial association rule is a rule which describes the implication of one a set of features by another set of features in spatial databases. For example, a rule associating the price range of the houses with nearby spatial features, like beaches, is a spatial association rule.

Thematic Maps: Thematic map is map primarily design to show a theme, a single spatial distribution or a pattern, using a specific map type. These maps show the distribution of features over limited geography areas [6-9]. Each map defines a partitioning of the area into a set of closed and disjoint regions; each includes all the points with the same feature value. Thematic maps present the spatial distribution of a single or a few attributes. This differs from general or reference maps where the main objective is to present the position of the object in relation to other spatial objects. Thematic maps may be used for discovering different rules. For example, we may want to look at temperature thematic map while analyzing the general weather pattern of a geographic region. There are two ways to represent thematic maps: Raster, and Vector. In the raster image form thematic maps have pixels associated with the attribute values. For example, a map may have the altitude of the spatial objects coded as the intensity of the pixel (or the color). In the vector representation, a spatial object is represented by its geometry, most commonly being the boundary representation along with the thematic attributes. For example, a park may be represented by the boundary points and corresponding elevation values.

III. SPATIAL DATA MINING TASKS

As shown in the table 1, spatial data mining tasks are generally an extension of data mining tasks in which spatial data and criteria are combined. These tasks aim to: (i) summarize data, (ii) find classification rules, (iii) make clusters of similar objects, (iv) find associations and dependencies to characterize data, and (v) detect deviations after looking for general trends. They are carried out using different methods, some of which are derived from statistics and others from the field of machine learning.

Spatial data summarization: The main goal is to describe data in a global way, which can be done in several ways. One involves extending statistical methods such as variance or factorial analysis to spatial structures. Another entails applying the generalization method to spatial data.

Statistical analysis of contiguous objects

Global autocorrelation: The most common way of summarizing a dataset is to apply elementary statistics, such as the calculation of average, variance, etc., and graphic tools like histograms and pie charts. New methods have been developed for measuring neighborhood dependency at a global level, such as local variance and local covariance, spatial auto-correlation by Geary, and Moran indices [10,11]. These methods are based on the notion of a contiguity matrix that represents the spatial relationships between objects. It should be noted that this contiguity can correspond to different spatial relationships, such as adjacency, a distance gap, and so on.

SDM Tasks	Statistics	Machine Learning
Summarization	Global autocorrelation Density analysis Smooth and contrast analysis Factorial analysis	Generalization Characteristics rules
Class identification	Spatial classification	Decision trees
Clustering	Point pattern analysis	Geometric clustering
Dependencies	Local autocorrelation Correspondence analysis	Association rules
Trends and deviations	Kriging	Trend rules

Table 1 comparisons

Density analysis: This method forms part of Exploratory Spatial Data Analysis (ESDA) which, contrary to the autocorrelation measure, does not require any knowledge about data. The idea is to estimate the density by computing the intensity of each small circle window on the space and then to visualize the point pattern. It could be described as a graphical method.

Smooth, contrast and factorial analysis: In density analysis, non-spatial properties are ignored. Geographic data analysis is usually concerned with both alphanumerical properties (called attributes) and spatial data. This requires two things: integrating spatial data with attributes in the analysis process, and using multidimensional data to analyze multiple attributes [12,13].

IV. SPATIAL DATABASE

Spatial database is the core of spatial infrastructure. The framework of spatial database has experienced three stages in the past several years - file system, relational database mixed with file system, spatial data engine and spatially enabled database. In today's spatial data processing environment, spatial data engine (SDE) and spatially enabled database (SEDBMS) are two mainstream approaches to manage spatial data (Figure 3). Essentially SDE is a middleware which is used as the exchange channel between front-end interface and back-end container. Relational databases are well-structured data containers compared with file system. Spatial data engine provides unified but proprietary interface for front-end user to access spatial data which are stored in different data containers. However, this approach may reduce some efficiency due to the mapping between middleware and back-end container. And it's difficult to integrate the query engine in SDE with back-end database query engine [14-20].

SEDBMS is used on Object-relational DBMS. ORDBMS provides the ability of abstract data types and operations definition. The spatial data can be stored and accessed like the traditional structured data. Users can add types and functions/methods of spatial data with standard extended SQL in which new data types and functionality are integrated as closely as possible into the DBMS. Traditional DBMS functionality such as indexing, query optimization, and transaction management are supported in a seamless fashion
For user-defined data types and functions [14-20].

The database interacts directly with user instead of through a SDE broker, which makes it easier to spatial databases operations and generally results in good performance. PostGIS, MySQL Spatial Extension, Ingres Spatial Object Library, Oracle Spatial, IBM DB2 Spatial Extender and Informix Spatial DataBlade are the examples of this spatial database. In spatially enabled database, spatial data and traditional schema data share same query interface, query engine and storage engine. So traditional QPE should be improved to process query workflow mixed with spatial and non-spatial predictions more efficiently[14-20].

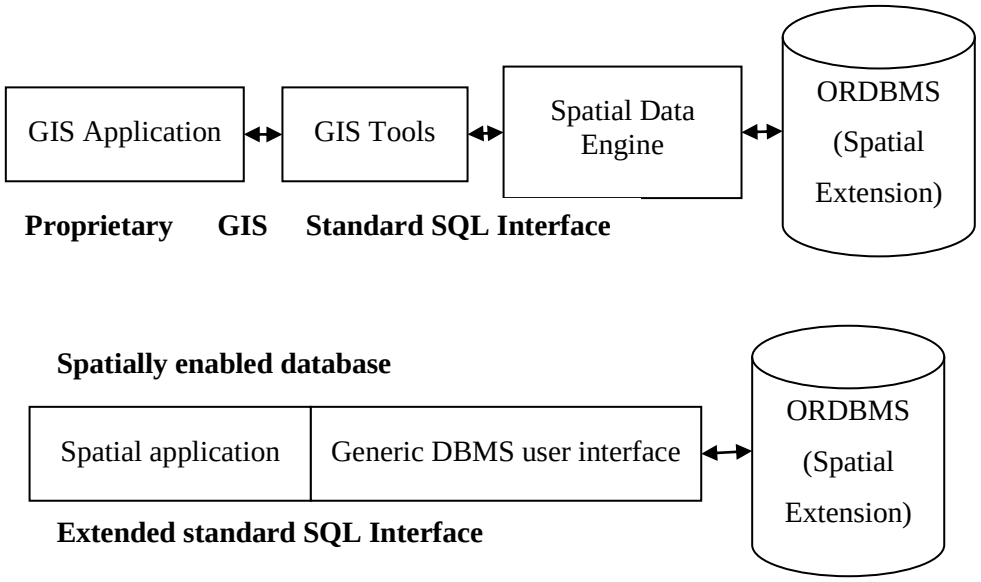

Fig 3 spatial data processing

V. QUERY EXAMPLES

One of the things that separates a GIS (or spatial) database from an "ordinary" database is the fact that it operates on entities that may have both spatial as well as alphanumerical attributes. Hence, queries to the database involve identification of these entities based on both the spatial and alphanumerical attributes (Ooi et al., 1989). In general there are two ways of achieving this, either to specifically design a new query language, or to extend an existing one with spatial operations [14-20].

Extension of existing query languages: Many attempts have been made using the latter approach, such as the SQL extension GEOQL (GEOgraphic Query Language) described by Ooi et al. (1989). Using GEOQL, the query finds all roads that intersect with roads that are adjacent to Monash University or RMIT" would be:

```
SELECT X.name
FROM road X, road Y, region
WHERE X intersects Y and Y is adjacent to Y and
(region.name = 'Montash University' or
region.name = 'RMIT')
```

As we see the SQL-form SELECT FROM WHERE is kept, and some geographical terms have been introduced (intersects, is adjacent to). Still, GEOQL is just one of many attempts to extend an existing (relational) query language. Egenhofer (1994) mentions GEO-QUEL (an extension of QUEL) and Query-By-Pictorial-Example (an extension of Query-By-Example)[14-20].

VI. CHALLENGES OF SPATIAL DATABASES

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

Comparison of classical data mining techniques with spatial data mining techniques: Relationships among spatial objects are often implicit. It is possible to materialize the implicit relationships into traditional data input columns and then apply classical data mining techniques [21-24].

Modeling semantically rich spatial properties, such as topology: The spatial relationship among locations in a spatial framework is often modeled via a contiguity matrix using a neighborhood relationship defined using adjacency and distance. However, 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.

Statistical interpretation models for spatial patterns: Spatial patterns, such as spatial outliers and co-location rules, are identified in the spatial data mining process using unsupervised learning methods. There is a need for an independent measure of the statistical significance of such spatial patterns. Spatial interest measures: The interest measures of patterns in spatial data mining are different from those in classical data mining, especially regarding the four important output patterns.

Effective visualization of spatial relationships: Visualization in spatial data mining is useful to identify interesting spatial patterns. The data inputs of spatial data mining have both spatial and non-spatial features. To facilitate the visualization of spatial relationships, research is needed on ways to represent both spatial and non-spatial features.

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.

Preprocessing spatial data: Spatial data mining techniques have been widely applied to the data in many application domains. However, research on the preprocessing of spatial data has lagged behind. Hence, 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.

VII. CONCLUSIONS

The main contribution of this paper is to define a set of basic operations for KDD in SDBS which should be supported by an SDBS. The definition of such a set of basic operations and their efficient support by an SDBS will speed up both, the development of new spatial KDD algorithms and their performance. Different methods of spatial databases have been outlined in this paper, which has shown that this process will help to develop better techniques.

REFERENCES

[1]. Talhofer, V., Hošková, Š., Kratochvíl, V., & Hofmann, A. (2009). Geospatial Data Quality. ICMT'09 - International conference on military technologies, pp. 570-578,2009.

[2]. Frawley W.J., Piatetsky-Shapiro G., Matheus J.: "Knowledge Discovery in Databases: An Overview", in: Knowledge Discovery in Databases, AAAI Press, Menlo Park, pp. 1-27, 1991.

[3]. Koperski K., Han J.: "Discovery of Spatial Association Rules in Geographic Information Databases", Proc. 4th Int. Symp. on Large Spatial Databases, Portland, ME, pp.47-66, 1995.

[4]. Shekhar, S.; Lu, C.; and Zhang, P.. A Unified Approach to Detecting Spatial Outliers. GeoInformatica 7(2), 2003.

[5]. Shekhar, S.; Schrater, P. R.; Vatsavai, R. R.; Wu, W.; and Chawla, S. Spatial Contextual Classification and Prediction Models for Mining Geospatial Data. IEEE Transaction on Multimedia 4(2), 2002.

[6]. Morimoto, Y. Mining Frequent Neighboring Class Sets in Spatial Databases. In Proc. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2001.

[7]. Zhang, P.; Huang, Y.; Shekhar, S.; and Kumar, V. Exploiting Spatial Autocorrelation to Efficiently Process Correlation-Based Similarity Queries. In Proc. of the 8th Intl. Symp. on Spatial and Temporal Databases, 2003.

[8]. M.Hemalatha.M; Naga Saranya.N. A Recent Survey on Knowledge Discovery in Spatial Data Mining, IJCI International Journal of Computer Science, Vol 8, Issue 3, No.2, may, 2011.

[9]. Shekhar, S., and Huang, Y. Co-location Rules Mining: A Summary of Results. In Proc. of the 7th Int'l Symp. on Spatial and Temporal Databases, 2001.

[10]. Geary R.C.: The contiguity ratio and statistical mapping, The incorporated Statistician, 5 (3), pp 115-145.

[11]. Moran P.A.P., The interpretation of statistical maps, Journal of the Royal Statistical Society, B: 10, pp 234-251.

[12]. Lebart, L. (1984) Correspondence analysis of graph structure. Bulletin technique du CESIA, Paris:2, 1-2, pp 5-19.

[13]. Benali, H., Escofier, B.: Analyse factorielle lissée et analyse factorielle des différences locales, Revue Statistique Appliquée, XXXVIII (2), pp 55-76, 1990

[14]. G. Brent Hall and Michael G. LeahyOpen Source Approaches in Spatial Data Handling, pp.105-129,2008

[15]. David W. Adler. DB2 Spatial Extender , "Spatial data within the RDBMS" in Proceedings of the 27th International Conference on Very Large Data Bases, pp.687-690, 2001

[16]. Ramsey P,PostGIS manual. Refractions Research Inc,2005

[17]. Graefe G., Query evaluation techniques for large databases. in ACM computing Surveys, 25(2), pp.73-170, 1993.

[18]. Yan Zhou, Spatial Data Dynamic Balancing Distribution Method for Parallel Spatial Database,2009

[19]. Guobin Li .Research on Optimized Spatial Data Query Algorithm in the Spatial Database.

[20]. Yan zhou .Spatial Data Declustering Method Considering Spatial Locality for Parallel Spatial Database.
[21]. Agrawal, R., and Srikant, R. Fast Algorithms for Mining Association Rules. In Proc. of Very Large Databases, 1994.
[22]. Jain, A., and Dubes, R. Algorithms for Clustering Data. Prentice Hall, 1988.
[23]. Quinlan, J. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, 1993.
[24]. Barnett, V., and Lewis, T. Outliers in Statistical Data. John Wiley, 3rd edition edition, 1994.