A WALK ON SPATIAL DATAMINING APPROACHES

[1]Dr. C.Anuradha, M.C.A., M.Phil.,SET,Ph.D.,
Assistant Professor and Head, Dept. of Computer Applications,
S.S.B.S. Tamil, Arts & Science College, Tamilnadu., India.

[2]R. Anandavally, M.Sc., M.Phil.,
Assistant Professor, Dept. Computer Science,
S.S.B.S. Tamil, Arts & Science College, Tamilnadu., India.

Abstract

Spatial Data mining is the discovery of interesting relationships and characteristics that may exist implicitly in spatial databases. A crucial challenge in spatial Data mining is the efficiency of spatial data mining algorithm due to the often huge amount of spatial data and the complexity of spatial data type's, spatial accessing methods. This research paper explores application of spatial data mining along with the techniques and the unique features that distinguish spatial data mining from classical data mining. Also it focuses on role of clusters in extracting spatial pattern and capturing intrinsic relationships between spatial and non-spatial data. Spatial data mining algorithms are examined to reorganize spatial databases to accommodate data semantics, as well as to achieve better performance. Major components and research needs in spatial data mining research are discussed. Finally it presents an area of Spatial Data mining where further research is needed.

Keywords: Geospatial, Clustering and Outliers, Association, Geographical Information System (GIS).

1. Introduction

Due to advancement in Database technology since the mid-1980s has been characterized by the popular adoption of relational technology and an upsurge of research and development happenings of different and powerful database systems. These promote the development of advanced data models such as object- oriented, object relational and deductive models. This explosively growing data creates a promising field, called data mining or knowledge mining from the data. Data mining represents the integration of several techniques from spatial data

analysis, information retrieval, recognition, image analysis, signal processing, computer graphics, bioinformatics, or psychology.

Due to the diversity of disciplines contributions to data mining, research is expected to generate a large variety of data mining systems. Our focus of this overview is on the major activities and methods of spatial data mining, i.e., discovery of interesting knowledge from spatial data. Spatial databases contain spatial related information. Spatial data mining techniques extract useful patterns from spatial data sets. Examples include geographic databases, very large scale integration or computer –aided design databases, and medical and satellite image databases. Important attribute of SDM is location. Spatial data in GIS is defined as elements that can be stored in a map, images, graph and tabular forms [1]. This research paper presents the major accomplishments of spatial data mining research, also focus on applications, various tasks and trends associated in spatial data mining are presented in detail. Furthermore, shows that spatial data mining is promising field with rich research is needed and many challenging issues. Finally, identifies the areas of spatial data mining where further research is needed.

2. Background and Motivations

The complexity of spatial data and intrinsic spatial relationships limits the usefulness of conventional data mining methods for extracting spatial patterns. Efficient tools for extracting information from geospatial 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) [2]. Extracting interesting and useful patterns from spatial data sets is more difficult than extracting corresponding patterns from traditional numeric and categorical data due to the complexity of spatial data types, spatial relationships, and spatial autocorrelation.

2.1 Literature Survey

A detailed discussion on the research work done by various researchers and review of the literature on the spatial data mining are carried out in this section. The researchers Hemlata et al. [3] had done a research work in geographical information system (GIS) and Spatial data mining

(SDM) in Big data. Researchers describes that increasing information from geo-databases has gained value using computational algorithms in the field of spatial data mining. Therefore, the

authors introduced a GIS and spatial data mining with the tools, algorithms approaches, research needs and issues, along with the role of spatial association rule mining in big data of GIS. Research work done by sophiya and saurabh sharma, had proposed a study related to K-means and Ward's algorithm with Honey Bee optimization is done for spatial data mining(SDM). The researchers reveals a new hybrid approach for clustering the data make use of FCM and gravitational search optimization [4].

Manjula et al. [5] had done survey in spatial data mining and show the traditional methods of clustering and association rules based on spatial data collected from large amount of spatial data bases. The article reviews latest mechanisms on spatial data mining, from spatial data generalization, to spatial data clustering, mining spatial association rules and also shows the research challenges presents in the spatial data mining. In the manuscript done by Venkateswara et al. had examined the issues and challenges related to spatiotemporal data mining tasks such as association rules, classification, clustering for discovering knowledge from spatiotemporal datasets are reviewed. For knowledge discovery, system functional requirements and database structure along with spatiotemporal data mining are also presented [6].

2.2 Moving Forward

The convergence of advances in location-aware computing, databases and knowledge discovery, and human interaction technologies, combined with the increasing quality and quantity of geospatial information, can transform our world. Diverse technological advances will be needed to attain that goal, and we must line up the talent and resources needed to achieve those advances. Only by maintaining the long-term view that geospatial information should be made accessible to everyone, everywhere, in appropriate and useful ways, can we exploit the full potential of geospatial information for enriching science and safeguarding society [7].

Specific features of geographical data that preclude the use of general purpose data mining algorithms are: i) rich data types (e.g., extended spatial objects) ii) implicit spatial

relationships among the variables, iii) observations that are not independent, and iv) spatial autocorrelation among the features in spatial data mining.

3. Spatial Data mining Applications

Spatial data mining has its application [8] in many fields. Few real- world applications are stated below.

- Business prospecting
- Hospital prospecting
- Spatial region-based classification or personalization
- Automobile insurance
- Mobile-commerce industry
- Local instability in traffic
- Migration of birds
- Autonomous navigation
- Fleet tracking
- Fishing control
- Pedestrian behavior analysis

4. Spatial Data mining Tasks

Data mining deals with the kind of pattern discovery to be mined from the spatial dataset. Here is the list of some of spatial data mining task.

4. 1 Spatial Classifications

Classification is about grouping data items into class according to their properties (attribute values). Spatial classification methods extend the general-purpose classification methods to consider not only attributes of the object be classified but also the attributes of neighboring objects and their spatial relations [9].

4.2 Spatial Association Rule

Association rule mining uses the concepts of support and confidence to identify interesting rules. The *support* is the probability of a record in the database satisfying the set of predicates contained in both the antecedent and consequent. The *confidence* is the probability that a record that contains the antecedent also contains the consequent.

Association rules generated from mining data at multiple levels of abstraction are called multilevel association rules. It can be mined effectively using concept hierarchies under a support-confident framework.

4.3 Spatial Clustering

Spatial clustering is a collection of spatial objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters [10]. Clustering is a very well-known technique and clustering algorithm to deal with the large geographical datasets. Clustering techniques organized in table 1.

Table 1: Spatial Cluster methods

Name of the Algorithms	Name of the Authors	Method Descriptions	Advantages	Limitations
Partitioning methods (K-means-medics, CLARANS)	Macqueen, Kaufman& Rousseeuw, Ng & Han	• They are made for minimizing the distance from the data objects to their distance centers [8, 10]. • Resulting is intra-cluster similarity is high.	• Work well in spherical shaped clusters in small Database. • It minimizes the square error Functions. • It enables the detection of Outliers.	• Difficult to find clusters in Complex shapes. • Lack of techniques to explore spatial Data structure. • Cannot find clusters in large Datasets.

Hierarchical methods (AGNES, DIANA,BIRCH, ROCK)	Kaufman& Rousseeuw, Zhang	• This algorithm fixed the membership of a data object once it has been allocated to a cluster. • It uses complex criteria for compressing and relocating data before merging clusters [11, 8].	• A single scan of dataset yields high Quality clusters. • Minimize the time required for I/O.	• Suffers from the fact that once a step is done, it can never be undone. • Difficulties in selection of merge or split which may lead to low quality clusters.
Density Based methods (DBSCAN, OPTICS, DENCLUE)	Tan H, MadH, Ester	This method is based on density of data points within a region to discover Clusters [10].	• It is used to filter out noise. • Effective in finding arbitrary Shaped clusters.	• Difficult to find real world, high dimensional data. • It has run time complexity.
Grid Based methods (STING, Wave cluster, CLIQUE)	Wang, Sheikhole-slami, Agrawal.	It quantizes the object space into a finite number of cell and operations performed in the grid structures.	• It has both efficiency and Cluster quality. • This approach is its fast processing time.	• Resulting Clusters are isothetic: this may lead to low quality and accuracy of Clusters [10].

4.4 Spatial Outlier Analysis

Spatial Outlier is an object whose non-spatial attribute value is significantly different from the values of its spatial neighbors. Outlier may be detected using statistical tests that assume a probability model for the data. For spatial outlier detection there exist three-step approaches to detect spatial outliers in large databases. These steps are clustering, checking spatial neighbors.

4.5 Spatial Prediction

Predictive mining tasks perform inference on the current data in order to make predictions. It is used to predict missing or unavailable numerical data values. Prediction may

refer to both numerical prediction and class label prediction [10]. Regression analysis is the methodology used for numerical prediction.

5. Trends in Spatial Data mining

Spatial Data mining Tasks aims at discovering various kinds of potentially useful and unknown patterns and Trends from spatial databases [9].

- Multidimensional analysis of spatial data.
- Spatial data characterization.
- Spatial topological Relationship discovery.
- Mining Spatial topological Relationship patterns.
- Spatial Neighborhood.
- Spatial Association Rule.
- Spatial data classification.
- Trend Prediction or Detection.
- Spatial Data Clustering.
- Spatial Outlier analysis.

6. Research Challenges and Legal Issues

Research investments will be required to develop dimensionality reduction methods that are scalable, robust, and nonlinear. Moreover, few current data mining algorithms can handle temporal dimensions, and even fewer can accommodate spatial objects other than points. Research must be directed at new techniques that will be capable of finding patterns in complex geospatial objects that move, change shape, evolve, and appear/disappear over time. Some of the overarching issues presented here.

- Policy Issues
- Precision Agriculture
- Accessible Location-Sensing Infrastructure
- Mobile Environments.
- Research at the Intersection of Information Technology and Geospatial Science

7. Future Trends

- Work further with geophysical data.
- Work in conjunction with other disciplines.
- Provide Data mining expertise.
- Work in Spatial & temporal multi-layered data sets from environmental Sciences.

8. Conclusion

Spatial Data mining is a promising research area with wide application of computational techniques for the analysis of spatial data. A grand challenge for science is to understand the human implications of global environmental change and to help society cope with those changes. This article presented the Overview and some of Spatial Clustering algorithms. In addition to that, discuss about tasks, trends and some of the research issues in spatial data mining. It requires new technology and sophisticated understand of the spatial problems associated with Spatial data. This geographic data mining or geographic knowledge is not new, but is developing and changing rapidly.

The aim is to assist in generation of hypothesis, which can be tested, about to find interesting and anomalous spatial patterns in large database. The intent is to support the humans in these tasks. A day of “Big Data” requires fresh thinking.

References

[1] Han, Jiawei, Micheline Kamber, Anthony KH Tung, "Spatial clustering methods in data mining", Geographic data mining and knowledge discovery, 2001, pp.188-217.

[2] Dr.M.Hemalatha, N.Naga Saranya, “A Recent Survey on Knowledge Discovery in Spatial Data Mining”, International Journal of Computer Science Issues, Vol. 8, Issue 3, No.2, May 2011.

[3] Hemlata Goyal, Chilka Sharma, Nisheeth Joshi, “ An Integrated approach of GIS and Spatial Data Mining in Big Data”, International Journal of Computer Applications, Vol. 169, No. 11, 2017,pp.1-6.

[4] Sophiya, Saurabh Sharma, "A Review on Hybrid Approach Based on K-Mean and Ward's Algorithm", International Journal for Research in Applied Science & engineering Technology, Vol. 5, Issue. 7, 2017, pp. 1075-1078.

[5] Manjula Aakunuri, G. Narasimha, Sudhakar Katherapaka, "Spatial Data Mining: A Recent Survey and New Discussions", International Journal of Computer science and Information Technologies, Vol.2, No.4, 2011, pp.1501-1504.

[6] Venkateswara Rao, Govardhan A, chalapati Rao K.V., "Spatiotemporal Data mining: Issues, Tasks and Applications", International Journal of Computer Science & Engineering Survey, Vol.3, No.1, 2012, pp.39-52.

[7] Majeed, Z. A., Parker, D.," Geographic Information System for Managing Survey Data to the Development of GIS-Ready Information", In Proceedings of the 3rd FIG Regional Conference, 2004.

[8] Dipika Kalyani, Setu Kumar Chaturwedi, "A Survey on Spatio-Temporal Data Mining", International Journal of Computer Science and Network Volume 1, Issues 4, August 2012.

[9] Sumathi, N., Geetha, R., Bama, S. S., "Spatial data mining techniques trends and its applications", Journal of Computer Applications, Vol.1, No.4, 2008, pp.28-30.

[10] Han, J., Pei, J., Kamber, M.,"Data mining: concepts and techniques", Elsevier, 2011.

[11] Koperski, K., Adhikary, J., Han, J., "Spatial data mining: progress and challenges survey paper", In Proc. ACM SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery, Canada,1996, pp. 1-10.