ISSN 2229-5518

An Overview of Fuzzy spatial Data mining in an Object oriented Environment

Sarita Tripathy, Prof. (Dr.) Laxman Sahoo

Abstract- Fuzzy spatial data mining technique has been developed to extract relationships describing relative positions of classes of objects from raster images. Demand has been increased for complex data in various applications. These complex data can be easily represented and manipulated by Object-oriented systems it is also capable of managing complicated and uncertain relationship existing among them. They are also much suitable for engineering and scientific applications. This paper gives an overview of fuzzy spatial data mining in an Object-oriented environment.

Index terms- fuzzy data mining, spatial data mining, fuzzy spatial data mining, uncertainty in data mining, fuzzy spatial databases, object oriented systems, fuzzy object oriented Database.

1. Introduction

Spatial data mining is a promising field of research with wide applications in GIS [1], medical imaging, remote sensing, and robot motion planning methods and so on. It encompasses various tasks, and for each task number of various methods are often available whether computational, statistical, visual, or some combination of them. Fuzzy spatial data mining is a new and exciting field that its bounds and potentials are yet to be defined. Object oriented databases [2] are based on object oriented programming paradigm where each entity is considered as an object. The fuzzy object oriented database (FOOD) model is one of the proposed models in literature to handle uncertainty in object oriented databases. Several kinds of fuzziness are dealt with in a FOOD model [3] including fuzziness at attribute level and between object and class and between class and super class relation. Spatial database systems [4] are full-fledged database systems which in addition to the functionality of standard database systems for alphanumeric data provide special support for the storage, retrieval, management and querying, that is objects in space. In particular, SDBS are used as the data management foundation of Geographical information systems. In literature the common consensus prevails that special data types are necessary to adequately model geometry and to efficiently represent geometric data in database systems. These data types are commonly denoted as spatial data types [5] such as point, line and region. We speak of spatial objects as instances of these data types; so far, the mapping of spatial phenomenon of real world leads almost exclusively to precisely define spatial objects. Spatial data modeling implicitly assumes that the positions of points, the locations and routes of lines, and the extent and hence the boundary of regions are precisely determined and universally recognized. This leads to exact object models. Examples are especially man-made spatial objects representing engineered artifacts(like monuments ,highways, buildings, bridges) and predominantly immaterial spatial objects exerting social control(like countries, districts and land parcels with their political administrative and cadastral boundaries).This kind of entities are denoted as crisp or determinate spatial objects[6]. Reminders of this paper organized as the second section gives an overview of Fuzzy spatial data mining. Third section gives a description of database oriented framework fore spatial data mining, fourth section is uncertainty in data mining, fifth section is fuzzy spatial databases, and sixth section is FOOD model and finally the conclusion

2. The Concept of Fuzzy spatial Data mining

Fuzzy spatial data mining is nothing but the combination of fuzzy data mining and spatial data mining which is described in the following sections:

2.1 Fuzzy data mining

Combining fuzzy logic with data mining processes results in fuzzy data mining techniques [7].With the proliferation of data, data mining tools are becoming available to meet the market demand for ways to find useful information within that data. One drawback to data mining, specifically data mining of spatial data, is representing vastly different data values and inferring missing data. This is especially

ISSN 2229-5518

evident in data mining applications that seek to find relationships between biological and environmental parameters. Current data mining approaches [8] that utilize neural networks, genetic algorithms, or statistical techniques do not inherently allow for such common data inadequacies. A methodology is needed that can properly represent, and process, data with a large amount of uncertainty. This drawback can be overcome by the use of fuzzy data mining methodology .An early and significant application of fuzzy sets [9] has been in pattern recognition, especially fuzzy clustering algorithms [10].Hence, much of the effort in fuzzy data mining has been by the use of fuzzy clustering and fuzzy set approaches in neural network and genetic algorithms. In fuzzy theory an important consideration is treatment of data from linguistic view point, from this an approach has been developed that uses linguistically quantified proposition to summarize the content of a database, by providing a general characterization of analyzed data. A common organization of data for data mining is the multidimensional data cube in data warehouse structures. Treating the data as a fuzzy object has provided another approach for knowledge discovery. Fuzzy data mining for generating association rules [11] has been considered by many researchers.

2.2 Spatial data mining

Spatial data mining[12] consists of extracting knowledge, spatial relationships and any other properties which are not explicitly stored in databases.SDM is used to find implicit regularities, relations between spatial data and/or non-spatial data. The specificity of SDM lies in its interaction in space. In effect, a geographical database constitutes a spatio-temporal continuum in which properties concerning a particular place are generally linked and explained in terms of the properties of its neighborhood.

We can thus see the great importance of spatial relationships in the analysis process. Temporal aspect of spatial data is also a central point but is rarely taken into account. Data mining methods are not suited to spatial data because they do not support location of data nor the implicit relationships between objects. Hence it is required to develop new methods including spatial relationships and spatial data handling. Calculating these spatial relationships is time consuming, and a huge volume of data is generated by encoding geometric location. Global performances will suffer from this complexity. Using GIS, the user can query spatial data and perform simple analytical tasks using programs and queries. However, GIS are not designed to perform complex data analysis or knowledge discovery. They do not provide generic methods for carrying out analysis and inferring rules.Neverthless, it seems necessary to integrate these existing methods and to extend them by incorporating spatial data miming methods.GIS methods are crucial for data access, spatial joins and graphical map display. Conventional data mining can only generate knowledge about alphanumerical properties. A primary goal of database research has been the incorporation of additional semantics into the data model. Traditionally data models have been good at representing precise, well structured data, however these methods have been found to be lacking in their ability to represent vague, uncertain or imprecise data. An approach to solving the problem has been to utilize fuzzy logic to extend the capabilities of the data model. The solutions to the problem of uncertainty management in databases have typically utilized fuzzy set theory. Fuzzy set theory holds that elements are associated in the fuzzy set with a degree of membership. In comparison, classical set theory holds that either an element is member of a set or not .i.e. there are no grades of membership. Using fuzzy set theory and its definitions, two separate approaches to the problem have emerged. The first is heterogeneous model, where the attributes may be possibility distributions, range values, fuzzy promote and fuzzy numbers. The second allows only homogenous data, such as fuzzy numbers and predicates, in attribute values. The latter approach utilizes the similarity relation as the basis for the representation and manipulation of data. The similarity relation determines the properties of any pair values existing in the domain. The values have the properties of symmetry, reflexivity and transitivity.

In this approach set valued attributes are permitted. Each fuzzy database domain has a similarity matrix associated with it. If the similarity between a pair of values is one, then the values are identical and if zero then they are completely dissimilar, any value in between indicates a level of similarity. Recently there has been considerable interest in object oriented database model, for complex engineering applications such as CASE, CAD/CAM etc. This has been motivated by the richness of the representational mechanisms provided by the model, specifically the ability of the data model to represent class-subclass relationships. Shared objects and multivalued data and are vital to the modeling of complex applications. The object oriented data model has been extended to represent models uncertainty.

Unlike other data model object oriented data model incorporates both behavior and structure. The incorporation of uncertainty affects the behavior of the data model.

2.3 Fuzzy spatial data mining

The research areas of data mining and knowledge discovery in databases have developed in response to the need for methods to automatically analyze this data and extract useful new knowledge implicit in the data. Remote sensing technologies have accounted for an explosive growth in the quantity of spatial data collected including satellite images and sonar images. Many data mining techniques that were originally developed for transaction data have been adapted for spatial data mining.Koperski, Han and Adhikary [13] define spatial data mining as "extraction of implicit knowledge", spatial relation or other patterns not explicitly stored in spatial databases. This definition assumes that the spatial data will be stored in a spatial database; however, much spatial data remains in raster form in flat files. The previous definition has been extended to cover all collections of spatial data, not just spatial databases. One aspect of data mining is the extraction of relationships among objects that are typical in certain data sets. Relative position is a prototypical vague concept. A statement such as "X is north of Y" may be true to a degree and is conveniently represented using fuzzy logic. Likewise as statements such as "objects of type A are typically north of objects of class B" is also vague and amenable to fuzzy logic representation. Fuzzy adaptation of association rule mining have been developed for mining descriptions of directional relationships between classes of objects in image data sets. Fuzzy spatial data mining [14] is nothing but the combination of fuzzy data mining and spatial data mining.

3. DATABASE ORIENTED FRAMEWORK FOR SPATIAL DATA MINING

A research group in Munich has developed a set of database primitives for mining in spatial databases that are sufficient to express most of the algorithms of spatial data mining [15] that can be efficiently supported by DBMS.They have found that the use of such database primitives enables the integration of spatial data mining with existing DBMS and speeds up the development of new spatial data mining algorithms. The database primitives are based on neighborhood graph and neighborhood path. Effective filter allow restriction of the search to such neighborhood path leading away from the starting object. Neighborhood indices materialize such neighborhood graphs to support efficient processing of the database primitives by DBMS.There are three basic types of 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 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. The relations are: A disjoint B,A meets B,A overlaps B,A equals B,A covers B,A covered-by B,A contains B,A inside B.A formal definition has been given Egenhofer(1991).Distance relations compare the distance of two objects with a given constant using one of the arithmetic comparison operators.

Fuzzy sets are sets whose elements have degrees of membership. Fuzzy sets were introduced simultaneously by lofti A.Zadeh and Dieter Klaua in 1965 as an extension of the classical notion of set. In classical set theory [16],the membership of elements in a set is assessed in binary terms according to a bivalent condition which is an element either belongs to or does not belongs to a set. By contrast fuzzy set theory permits the gradual assessment of the membership of elements in a set; this is described with the aid of a membership function valued in the real unit interval [0, 1].Fuzzy sets generalize classical sets, since the indicator functions of classical sets are special cases of the membership functions of fuzzy sets, if the latter only take values 0 or 1.in fuzzy set theory, classical bivalent set are usually crisp sets. The fuzzy set theory can be used in wide range of domains in which information is incomplete or imprecise, such as bioinformatics.

4. UNCERTAINITY IN DATAMINING

Data is often associated with uncertainty [17] because of measurement inaccuracy, sampling discrepancy, outdated data sources, or other errors. This is especially true for applications that require interaction with the physical world, such as location based services and sensor monitoring. For example, in the scenario of moving objects (such as vehicles or people), it is impossible for the database to track the exact locations of all objects at all time instants. Therefore, the location of each object is associated with uncertainty between update. These various sources of uncertainty have to be considered in order to produce accurate query and mining results .In recent years; there has been much research on the management of uncertain data in databases, such as the representation of uncertainty in databases and querying data with uncertainty. However, little research work has addressed the issue of mining uncertain data. It is noted that with uncertainty the data values are no longer atomic. To apply traditional data mining techniques, uncertain data has to be summarized into atomic values. Taking moving object application as an example again, the location of an object can be summarized either by its last recorded location or by an expected location (if probability distribution of an object's location is taken into account).Unfortunately the discrepancy in the summarized recorded values and actual values could seriously affect the mining results. If we solely rely on the recorded values, many objects could possibly be put into wrong clusters. Even worse each member of a cluster would change the cluster centroids, thus resulting in more errors.

5. Fuzzy object-oriented DATABASE (FOOD) Model

Many researchers have pointed out that OO database may be better choice for handling spatial data rather than traditional relational or extended relational models. Complex objects might have uncertain information, which an object-oriented data model should consider. FOOD model is a similarity-based data model. The study extends this model for better representation of uncertainty. Fuzzy object oriented database are tested as much suitable to represent and manipulate the spatial data and also fuzzy spatial data.

6.1 Object- oriented transactions

An object-oriented transaction for example includes one or more purchased items, each of which is represented as an object or an instance. Each instance inherits its characteristics from a superior object, called class, which defines the basic structure of objects with common properties, including attributes are used to represent the characteristics of a class, and the methods are used to implement the operations and functions of a class. An OODBMS is a system upon which the database is implemented. It is possible (but not optimal) to model using an object-oriented methodology and implement in, for example a relational DBMS.Of course, it is most desirable to use an OODBMS which can naturally implement all the constructs of the data model.However,owing to the newness of the technology,OODBMS are only now emerging as viable systems.

Recent description of some of the most innovative of such systems is given in KIM and lochovsky (1989).Such systems have important impact on GIS technology. For example most OODBMS support version control, where the system can generate multiple different versions of an object, may be corresponding to different time- slices. This would be a natural implementation of spatially referenced data (e.g. census data) where spatial boundaries of object oriented data modeling for spatial databases can change. Object oriented databases have become quite popular for many reasons. Classes and inheritance allow for code reuse through specialization and generalization, and encapsulation packages the data and methods that act on the data together in an object. Objects can be defined to represent very complex data structure and to model relationships in the data, as is often the case with spatial data. Object modeling helps in understanding the requirements of an enterprise, and object-oriented techniques lead to high quality systems that are easy to modify and maintain. Because many newer applications involving CAD/CAM, multimedia and GIS are not suitable for standard relational database model, object –oriented databases may be developed to meet the needs of these more complex applications.

7. CONCLUSION

Object oriented databases handles complex and uncertain data but it is unable to handle very complex data's which increases with the introduction of spatial data. This can be handled more efficiently by introducing fuzziness into the object oriented data model. Reasoning inexact information extensively exists in data and knowledge intensive applications and fuzzy techniques plays vital role to handle such type of information in modeling at conceptual and logical level, query and data processing, indexing and implementations of the next generation database systems. Fuzzy object oriented databases are the natural fit for many engineering and scientific applications suffering from the representation and manipulation of inexact information precisely. This paper suggests that fuzzy spatial data mining in a fuzzy object oriented data model can be very much advantageous and can overcome many drawbacks. From the study it can be hoped that the uncertainty in spatial data can be further reduced by introducing fuzziness in the object oriented data model.

REFERENCES:

[1] Ferdinand Di Martino, Vincenzo Loia, Salvatore Sessa, and Michele Giodano," Fuzzy Reliability Analysis in the implementation of Geographical Information Systems.IEEE Transactions, 2002,pp - 980-984.

[2] Michael F.Worboys,Hilary M.hearnshaw,and Davidj.maguire,"Object-Oriented Data Modelling for Spatial Database",International Journal of Geographic Information Systems,Vol-4,Issue 4, 1990, pp.369-383.

[3] Adnan Yazici,Cagri ince,Murat K oyuncu,"FOOD Index:A Multidimensional index structure for Similarity-based Fuzzy Object Oriented Database Models',IEEE Transactions on Fuzzy Systems,Vol.16,Issue 4,2008,pp.942-957.

[4] Ralf hartmut Guting,"An introduction to Spatial Database Systems",VLDB Journal,Vol.3,Issue 4,1994,pp.1-32.

[5] Markus Schneider,"Spatial data Types: Conceptual Foundation for the Design and Implementation of Spatial Database Systems and GIS",pp.1-71.

[6] Markus Schneider,"Finite Resolution crisp and Fuzzy Spatial Objects", pp.1-15.

[7] Jiawei Han.hong Cheng.Dong Xin.Xifeng Yan,"Frequent pattern mining:current status and future directions", Springer Science plus Business Media",2007,pp 56-86.

[8] Shivangi,Sengar,Shakshipriya and Urvashi Khatuja,"Fuzzy Set Approaches to Data Mining of Association Rule", International Journal of Computer Science and Informatics,Vol.2,Issue.4,2012,pp.33-36.

[9] Hemanta K.Baruah,"The Theory of Fuzzy sets:Beliefs and Realities", International Journal of Energy, Information and Communications,Vol.2,Issue.2,2001,pp.1-22

[10] K.M.Bataineh,M.Najia,M.Saqera,"Comparision Study Between Various Fuzzy Clustering Algorithms",Jordan Journal of Mechanical and Industrial Engineering,Vol 5,Issue.4,2011,pp.335-343.]

[11] Zongyao Sha,.Xiaolei Li,"Mining Local Association Pattern From Spatial Data Set",7th International Conference on Fuzzy Systems and Knowledge Discovery ,2010,pp. 1455-1459.

[12] Theres a Beaubouef,Frederick E Pehy."A Rough set foundation of spatial data mining involving vague regions",IEEE Transactions,2002,pp 767-772.

[13] George Brannon Smith Susan M.Bridges,"Fuzzy Spatial Data mining", IEEE Transactions, 2002, pp.1-7.

[14] Han J and Kamber," Data mining; Concepts and Techniques", Book.

[15] Arvind Sharma, HS Jat, RK Gupta," A Survey of Spatial Data mining Approaches: Algorithms and Architecture", International Journal of Computer Technology and Electronics Communication, pp. 15-22.

[16] Gao QS,Gao XY,Hu Y,"A New Fuzzy Set Theory Satisfying all Fuzzy Set Formulas", Journal of Computer Science and Technology,Vol 24,Issue no. 4,2009,pp-798-804.

[17] Hanning Yuan, Shuliang Wang," Factors Causing Uncertainties in Spatial Data mining",International Archives of Photogammetry,Remote sensing and Spatial Information Sciences,Vol 34,pp-1-4.