International Journal of Information Technology and Management

Vol. VI, Issue No. I, February-2014, ISSN 2249-4510

AN EVALUATION UPON CONCEPT AND TECHNIQUES OF SPATIAL DATA MINING

AN INTERNATIONALLY INDEXED PEER REVIEWED & REFEREED JOURNAL

www.ignited.in

International Journal of Information Technology and Management
Vol. VI, Issue No. I, February-2014, ISSN 2249-4510

An Evaluation upon Concept and Techniques of Spatial Data Mining

Dr. Shailendra Singh Sikarwar[1] Mahesh Bansal[2]

[1]Assistant Professor, P. G. V. College, Gwalior

[2]Assistant Professor, P. G. V. College, Gwalior

Abstract – A growing attention has been paid to spatial data mining and knowledge discovery (SDMKD). This paper presents the principles of SDMKD, proposes three new techniques, and gives their applicability and examples. First, the motivation of SDMKD is briefed. Second, the intension and extension of SDMKD concept are presented. Third, three new techniques are proposed in this section, i.e. SDMKD-based image classification that integrates spatial inductive learning from GIS database and Bayesian classification, cloud model that integrates randomness and fuzziness, data field that radiate the energy of observed data to the universe discourse. Fourth, applicability and examples are studied on three cases. The first is remote sensing classification, the second is landslide-monitoring data mining, and the third is uncertain reasoning.

Spatial data mining algorithms intensely rely on upon the proficient processing of neighborhood relations since the neighbors of numerous items must be researched in a solitary run of a normal calculation. Along these lines, giving general thoughts behind neighborhood relations and additionally an effective usage of these notions will permit a tight joining of spatial data mining algorithms with a spatial database management system. This will speed up both, the improvement and the execution of spatial data mining algorithms. In this paper, we characterize neighborhood graphs and ways and a little set of database primitives for their control. We demonstrate that normal spatial data mining algorithms are overall backed by the proposed fundamental operations. For discovering critical spatial examples, just certain classes of ways "heading endlessly" from a beginning item are significant.

This paper highlights recent theoretical and applied research in spatial data mining and knowledge discovery. We first briefly review the literature on several common spatial data-mining tasks, including spatial classification and prediction; spatial association rule mining; spatial cluster analysis; and geovisualization. The articles included in this special issue contribute to spatial data mining research by developing new techniques for point pattern analysis, prediction in space–time data, and analysis of moving object data, as well as by demonstrating applications of genetic algorithms for optimization in the context of image classification and spatial interpolation.

----------♦----------

INTRODUCTION

Data mining is the process of identifying fascinating and conceivably advantageous examples of information installed in substantial databases. The mining analogy is intended to pass on an impression that examples are chunks of valuable information covered up inside vast databases holding up to be identified. Data mining has been rapidly grasped by the business planet as a method for outfitting information from the a lot of data that enterprises have gathered and carefully archived through the years.

In the event that data mining is about concentrating examples from substantial databases, then the biggest databases have an in number spatial part. Case in point, the Earth Observation Satellites, which are systematically mapping the whole surface of the earth, gather in the vicinity of one terabyte of data consistently. Other expansive spatial databases might be the U.s. enumeration, and the climate and atmosphere databases. The necessities of mining spatial databases are not the same as those of mining established social databases. Specifically, the thought of spatial autocorrelation that comparative items have a tendency to bunch in geographic space, is key to spatial data mining.

The complete data-mining process is a blend of numerous sub processes which are worthy of study in their right.. Some significant sub processes are data extraction and data cleaning, characteristic

determination, calculation outline and tuning, and the analysis of the yield when the calculation is connected to the data. For spatial data, the issue of scale the level of aggregation at which the data are constantly examined, is additionally exceptionally critical. It. is well known in spatial analysis that indistinguishable investigations at. diverse levels of scale can at times lead to contradictory results. Our center in this part is constrained to the configuration of data-mining algorithms.

Spatial data are more complex, more changeable and bigger that common affair datasets. Spatial dimension means each item of data has a spatial reference (Haining, 2003) where each entity occurs on the continuous surface, or where the spatial-referenced relationship exists between two neighbor entities. Spatial data includes not only positional data and attribute data, but also spatial relationships among spatial entities. Moreover, spatial data structure is more complex than the tables in ordinary relational database. Besides tabular data, there are vector and raster graphic data in spatial database. And the features of graphic data are not explicitly stored in the database. At the same time, contemporary GIS have only basic analysis functionalities, the results of which are explicit. And it is under the assumption of dependency and on the basis of the sampled data that geostatistics estimates at unsampled locations or make a map of the attribute. Because the discovered spatial knowledge can support and improve spatial data-referenced decision-making, a growing attention has been paid to the study, development and application of SDMKD (Han, Kamber, 2001; Miller, Han, 2001; Li et al, 2001; 2002).

Spatial Database Systems (SDBS) are database systems for the management of spatial data. Both, the number and the extent of spatial databases are quickly developing in provisions for example, geo marketing, movement control and ecological studies. This development by far surpasses human limits to break down the databases with a specific end goal to discover verifiable regularities, manages or bunches covered up in the data. Thusly, mechanized knowledge discovery gets to be more essential in spatial databases. Knowledge discovery in databases (KDD) is the non-minor extraction of implied, at one time obscure, and possibly convenient information from databases.

The computerization of numerous business and government transactions and the developments in experimental data gathering devices give us an enormous and constantly expanding measure of data. This hazardous development of databases has far outpaced the human capability to decipher this data, making an dire necessity for new strategies and instruments that backing the human in converting the data into advantageous information and knowledge. Knowledge discovery in databases (KDD) has been characterized as the non-insignificant process of identifying substantial, novel, and possibly functional, and at last justifiable designs from data. The process of KDD is intelligent and iterative, including some steps, for example, the accompanying ones:

- Selection: selecting a subset of all qualities and a subset of all data from which the knowledge ought to be identified.
- Data lessening: utilizing dimensionality diminishment or conversion systems to lessen the viable number of ascribes to be acknowledged.
- Data mining: the provision of suitable algorithms that, under satisfactory computational proficiency confinements, prepare a specific list of examples over the data.
- Evaluation: translating and assessing the ran across examples regarding their functionality in the given provision.

Spatial Database Systems (SDBS) are database systems for the management of spatial data. To discover verifiable regularities, administers or examples covered up in expansive spatial databases, e.g. for geo-promoting, activity control or ecological studies, spatial data mining algorithms are exceptionally significant for a diagram of spatial data mining).

Most existing data mining algorithms run on divide and uniquely ready records, yet mixing them with a database management system (DBMS) has the accompanying preferences. Excess space and potential inconsistencies might be kept away from. Besides, business database systems offer different list structures to help distinctive sorts of database questions. This practicality can be utilized without additional execution exertion to speed-up the execution of data mining algorithms (which, as a rule, need to perform numerous database inquiries). Like the social standard dialect SQL, the utilization of standard primitives will speed-up the advancement of new data mining algorithms also will likewise make them more portable.

SPATIAL DATA MINING

Spatial data mining is a special kind of data mining. Tlie main difference between data mining and spatial data mining is that in spatial data mining tasks we use not only non-spatial attributes (as it is usual in data mining in non-spatial data), but also spatial attributes.

Spatial data mining is the process of discovering interesting and previously un-known, but potentially 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. Specific features of geographical data that preclude the use of general purpose data mining algorithms are:

- Rich data types (e.g., extended spatial objects)
- Implicit spatial relationships among the variables
- Observations that are not independent, and
- Spatial autocorrelation among the features.

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.

COMMON SPATIAL DATA-MINING TASKS

Spatial data mining is a growing research field that is still at a very early stage. During the last decade, due to the widespread applications of GPS technology, web-based spatial data sharing and mapping, high-resolution remote sensing, and location-based services, more and more research domains have created or gained access to high-quality geographic data to incorporate spatial information and analysis in various studies, such as social analysis (Spielman & Thill, 2008) and business applications (Brimicombe,

2007). Besides the research domain, private industries and the general public also have enormous interest in both contributing geographic data and using the vast data resources for various application needs. Therefore, it is well anticipated that more and more new uses of spatial data and novel spatial data mining approaches will be developed in the coming years. Although we attempt to present an overview of common spatial data mining methods in this section, readers should be aware that spatial data mining is a new and exciting field that its bounds and potentials are yet to be defined.

Spatial data mining encompasses various tasks and, for each task, a number of different methods are often available, whether computational, statistical, visual, or some combination of them.

Here we only briefly introduce a selected set of tasks and related methods, including classification (supervised classification), association rule mining, clustering (unsupervised classification), and multivariate revisualization.

Spatial classification and prediction - Classification is about grouping data items into classes (categories) according to their properties (attribute values). Classification is also called supervised classification, as opposed to the unsupervised classification (clustering). ‘‘Supervised” classification needs a training dataset to train (or configure) the classification model, a validation dataset to validate (or optimize) the configuration, and a test dataset to evaluate the performance of the trained model.

Classification methods include, for example, decision trees, artificial neural networks (ANN), maximum likelihood estimation (MLE), linear discriminant function (LDF), support vector machines (SVM), nearest neighbor methods and case-based reasoning (CBR).

Spatial classification methods extend the general-purpose classification methods to consider not only attributes of the object to be classified but also the attributes of neighboring objects and their spatial relations (Ester, Kriegel, & Sander, 1997; Koperski, Han, & Stefanovic, 1998). A visual approach for spatial classification was introduced in (Andrienko & Andrienko, 1999), where the decision tree derived with the traditional algorithm C4.5 (Quinlan, 1993) is combined with map visualization to reveal spatial patterns of the classification rules. Decision tree induction has also been used to analyze and predict spatial choice behaviors (Thill & Wheelerm, 2000). Artificial neural networks (ANN) have been used for a broad variety of problems in spatial analysis (Fischer, 1998; Fischer, Reismann and Hlavackova-Schindler, 2003; Gopal, Liu and Woodcock, 2001; Yao & Thill, 2007). Remote sensing is one of the major areas that commonly use classification methods to classify image pixels into labeled categories (for example, Cleve, Kelly, Kearns, & Morltz, 2008).

Spatial clustering, regionalization and point pattern analysis - Cluster analysis is widely used for data analysis, which organizes a set of data items into groups (or clusters) so that items in the same group are similar to each other and different from those in other groups (Gordon, 1996; Jain & Dubes, 1988; Jain, Murty, & Flynn, 1999). Many different clustering methods have been developed in various research fields such as statistics, pattern recognition, data mining, machine learning, and spatial analysis.

Clustering methods can be broadly classified into two groups: partitioning clustering and hierarchical clustering. Partitioning clustering methods, such as K-means and self-organizing map (SOM) (Kohonen, 2001), divide a set of data items into a number of non-overlapping clusters. A data item is assigned to the ‘‘closest” cluster based on a proximity or

dissimilarity measure. Hierarchical clustering, on the other hand, organizes data items into a hierarchy with a sequence of nested partitions or groupings (Jain & Dubes, 1988). Commonly-used hierarchical clustering methods include the Ward's method (Ward, 1963), single-linkage clustering, average-linkage clustering, and complete-linkage clustering (Gordon, 1996; Jain & Dubes, 1988).

To consider spatial information in clustering, three types of clustering analysis have been studied, including spatial clustering (i.e., clustering of spatial points), regionalization (i.e., clustering with geographic contiguity constraints), and point pattern analysis (i.e., hot spot detection with spatial scan statistics). For the first type, spatial clustering, the similarity between data points or clusters is defined with spatial properties (such as locations and distances).

Spatial clustering methods can be partitioning or hierarchical, density-based, or grid-based. Readers are referred to (Han, Kamber, & Tung, 2001) for a comprehensive review of various spatial clustering methods.

ENCOURAGING SPATIAL DATA ALINING

We now present an illustration which will be utilized all around this section to outline the distinctive ideas in spatial data mining. We are given data in the vicinity of two wetlands on the shores of Lake Erie in Ohio, USA, to anticipate the spatial appropriation of a bog reproducing winged animal, the red-winged blackbird (Agelaius phoeniceus). The names of the wetlands are Darr and Stubble, and the data was gathered from April to June in two progressive years, 1995 and 1996.

An uniform lattice was forced on the two wetlands, and diverse sorts of estimations were recorded at each one unit or pixel. The span of every pixel was five meters. The qualities of seven qualities were recorded at each one unit, obviously area knowledge is pivotal in choosing which traits are imperative and which are definitely not.

For instance, Vegetation Durability was picked over Vegetation Species in light of the fact that specific knowledge about the settling propensities of the red-winged blackbird prescribed that the decision of home area is more reliant on the plant structure and its imperviousness to wind furthermore wave activity than on the plant species.

Measures of Spatial Form : Mean focus is the normal area, figured as the mean of X and mean of Y directions. The mean focus is otherwise called the inside of gravity of a spatial dispersion. Regularly the weighted mean focus is proper measure for some spatial provisions, for e.g., focal point of populace. The weighted mean focus is registered as the degree between the aggregate of the directions of each one focus multiplied by its weight (e.g., number of individuals in piece) furthermore the total of the weights. The measure focus is utilized as a part of some structures. It might be utilized to streamline complex items (e.g., to stay away from space prerequisites and unpredictability of digitations of verges, a geographic item could be spoken to by its focus), or for distinguishing the best area for an arranged movement (e.g. a dissemination focus ought to be spotted a main issue with the goal that make a trip to it is minimized).

Scattering is a measure of the spread of an appropriation around its focus. Regularly utilized measures of scattering and variability are reach, standard deviation, change and coefficient, of difference. Scattering measures for geographical disseminations are regularly figured as the summation over the proportion of the weight of geographic articles and the closeness between them. Shape is multi-dimensional, and there is no single measure to catch the sum of the measurements of the shape. A hefty portion of shape measures are dependent upon correlation of the shape's border with that of a ring of the same area.

The Data-Mining Trinity : Data mining is a without a doubt multidisciplinary area, and there are numerous novel methods for concentrating designs from data. Still, if one were to name data-mining systems, then the three most non-controversial marks might be arrangement, grouping, and cooperation standards. When we depict each of these classes in portion, we exhibit some illustrative illustrations where these systems might be connected.

The objective of characterization is to gauge the worth of a trait of a connection dependent upon the worth of the connection's different characteristics. Numerous issues could be communicated as characterization issues. Case in point, determining the areas of homes in a wetland based upon the worth of different characteristics (vegetation toughness, water profundity) is a grouping issue frequently additionally called the area expectation issue. Also, foreseeing where to need problem areas in wrongdoing action might be given a part as an area forecast issue. Retailers basically settle a area expectation issue when they choose an area for another store. The well- known representation in land, "Location is everything," is a prevalent indication of this issue.

ALGORITHMS FOR SPATIAL DATA MINING

To help our claim that the expressivity of our spatial data mining primitives is satisfactory, we exhibit how ordinary spatial data mining algorithms could be combined with a spatial DBMS by utilizing the database primitives presented as a part of area 2.

Spatial Clustering : Clustering is the errand of assembling the objects of a database into genuine subclasses (that is, bunches) with the goal that the parts of a bunch are as comparable as could be

expected under the circumstances though the parts of distinctive bunches contrast however much as could be expected from one another. Provisions of bunching in spatial databases are, e.g., the discovery of seismic blames by assembling the passages of a tremor inventory or the formation of topical maps in geographic information systems by grouping characteristic spaces.

Distinctive sorts of spatial grouping algorithms have been proposed, e.g. k-medoid grouping algorithms for example, CLARANS. This is a case of a worldwide clustering calculation (where a change of a solitary database item may impact all groups) which can't make utilization of our database primitives in a common manner. Then again, the essential thought of a solitary output calculation is to assembly neighboring objects of the database into groups dependent upon a nearby group condition performing stand out output through the database. Single output grouping algorithms are productive if the recovery of the neighborhood of an item might be effectively performed by the SDBS. Note that nearby group conditions are generally underpinned by our database primitives, specifically by the neighbors operation on a suitable neighborhood chart. The algorithmic composition of single output grouping is delineated in figure 1. Diverse bunch conditions yield distinctive thoughts of a group and diverse grouping algorithms.

For example, *GDBSCAN* (*Generalized Density Based Spatial Clustering of Applications with Noise*) relies on a density-based notion of clusters. The key idea of a density based cluster is that for each point of a cluster its *Eps*-neighborhood for some given *Eps* > 0 has to contain at least a minimum number of points, i.e. the "density" in the *Eps*-neighborhood of points has to exceed some threshold. This idea of "density-based clusters" can be generalized in two important ways. First, any notion of a neighborhood can be used instead of an *Eps*-neighborhood if the definition of the neighborhood is based on a binary predicate which is symmetric and reflexive.

Second, instead of simply counting the objects in a neighborhood of an object other measures to define the "cardinality" of that neighborhood can be used as well. Whereas a distance-based neighborhood is a natural notion of a neighborhood for point objects, it may be more appropriate to use topological relations such as *intersects* or *meets* to cluster spatially extended objects such as a set of polygons of largely differing sizes. for a detailed discussion of suitable neighborhood relations for different applications.

```
SingleScanClustering(Database db; NRelation rel)
    set Graph to create_NGraph (db, rel) ;
    initialize a set CurrentObjects as empty;
    for each node O in g do
        if O is not yet member of some cluster then
            create a new cluster C;
            insert O into CurrentObjects;
            while CurrentObjects not empty do
                remove the first element of CurrentObjects as O;
                set Neighbors to neighbors (Graph, O, TRUE) ;
                if Neighbors satisfy the cluster condition do
                    add O to cluster C;
                    add Neighbors to CurrentObjects;
end SingleScanClustering;
```

Figure 1. Schema of single scan clustering algorithms

Spatial Characterization : The task of *characterization* is to find a compact description for a selected subset of the database. In this section, we discuss the task of characterization in the context of spatial databases and review two relevant methods.

The algorithm presented in to find spatial association rules consists of 5 steps. Step 2 (coarse spatial computation) and step 4 (refined spatial computation) involve spatial aspects of the objects and are briefly examined in the following. Step 2 computes spatial joins of the object type to be characterized (such as town) with each of the other specified object types (such as water, road, boundary or mine) using a neighborhood relation (such as close-to). For each of the candidates obtained from step 2 (and which passed an additional filter-step 3), the exact spatial relation, for example *overlap*, is determined in step 4. Finally, a relation such as the one depicted in figure 2 results which is the input for the final step of rule generation. It is easy to see that the spatial steps 2 and 4 of this algorithm can be well supported by the neighbors operation on a suitable neighborhood graph.

Town	Water	Road	Boundary
Saanich	<meet, J.FucaStrait>	<overlap,highway1>, <close-to,highway17>	<close-to,US>
PrinceGeorge		<overlap, highway97>	
Petincton	<meet,OkanaganLake>	<overlap, highway97>	<close-to,US>
...	...	...	...

Figure 2. Input for the step of rule generation.

[EFKS 98] introduces the following definition of spatial characterization with respect to a database and a set of target objects which is a subset of the given database. A *spatial characterization* is a

description of the spatial and non-spatial properties which are typical for the target objects but not for the whole database. The relative frequencies of the non-spatial attribute values and the relative frequencies of the different object types are used as the interesting properties. For instance, different object types in a geographic database are communities, mountains, lakes, highways, railroads etc. To obtain a *spatial* characterization, not only the properties of the target objects, but also the properties of their neighbors (up to a given maximum number of edges in the relevant neighborhood graph) are considered.

SPATIAL DATA MINING: EXPLORATORY DATA ANALYSIS TECHNIQUES

Spatial data analysis comprises a broad spectrum of techniques that deals with both spatial and non-spatial characteristics of spatial objects. Exploratory techniques allow to investigate first and second order effects of the data. First order variation informs about large scale global trend of phenomena which is spatially distributed through the area. Hie second order variation defines dependence between observations.

- Global Autocorrelation (Moran's I): Moran's I is a measure of global spatial autocorrelation. Global or local autocorrelation reveal feature similarity based location and attribute values to explore the pattern whether it is clustered, dispersed, or random.

- Hot Spots (Getis-Ord): Tlie G-statistic is often used to identify whether hot spots or cold spots exist based oil so-called distance statistics. Hot spots are regions that stand out compared with the overall behaviour prevalent in the space. Hot spots can be detected by visualizing the distribution in format of choropleth or isarithmic
maps.

- Local Autocorrelation (Anselin's Local Moran I): The local Moran statistic is used as a local indicator of spatial association which is calculated for individual zones around each observation within a defined neighbourhood to identify similar or different pattern in nearby. Because the distribution of the statistic is not known,
high positive or high negative standardized scores of Ii are taken as indicators of similarity or dissimilarity respectively.

- Density (Kernel): Kernel density estimation is a nonparametric unsupervised learning procedure (classifier). Kernel, k is bivarate probability density function which is symmetric around the origin.

CONCLUSION

Tlie spatial data milling is newly arisen area when computer technique, database applied technique and management decision support techniques etc. have been developed at certain stage. The spatial data mining gathered productions that come from machine learning, pattern recognition, database, statistics, artificial intelligence and management information system etc. Different theories, put forward the different methods of spatial data mining, such as methods in statistics, proof theories, rule inductive, association rules,
cluster analysis, spatial analysis, fuzzy sets, cloud theories, rough sets, neural network, decision tree and spatial data mining technique based on information entropy etc. Spatial data mining, has established itself as a complete and potential area of research.

Data mining is a quickly developing area which lies at the crossing point of database management., statistics and manufactured clever. Data mining furnishes self-loader procedures for running across startling examples in quite substantial amounts of data. Spatial data mining is a specialty area inside data mining for the fast analysis of spatial data. Spatial data mining has can possibly impact real deductive tests incorporating worldwide environmental change and genomics.

The recognizing normal for spatial data mining could be flawlessly compressed by the in the first place law of geology: All things are identified yet close-by things are more identified than removed things. The suggestion of this articulation is that the standard suspicion of autonomy also indistinguishably disseminated (iid) arbitrary variables, which describe established data mining, is not relevant for the mining of spatial data. Spatial statisticians have instituted the saying spatial-autocorrelation to catch this property of spatial data.

We contend that these operations are sufficient for KDD algorithms recognizing spatial neighborhood relations by displaying the usage of four run of the mill spatial KDD algorithms dependent upon the proposed operations. Two of these algorithms are well-known from literary works, the other two algorithms are new and are imperative commitments to elucidate the contrasts between KDD in social and in spatial databases. Besides, the productive backing of operations on vast neighborhood graphs and on extensive sets of neighborhood ways by the SDBS is examined. Neighborhood lists are presented to emerge chose neighborhood graphs to speed up the processing of the proposed operations.

REFERENCES

- Agrawal R., Imielinski T., Swami A.: *"Database Mining: A Performance Perspective"*, IEEE Transactions on Knowledge and Data Engineering, Vol.5, No.6, 1993, pp. 914-925.

- Anselin, L., Syabri, I., & Kho, Y. (2006). GeoDa: An introduction to spatial data analysis. Geographical Analysis, 38(1), 5–22.
- Bill, Fritsch: "*Fundamentals of Geographical Information Systems: Hardware, Software and Data*" (in German), Wichmann Publishing, Heidelberg, Germany, 1991.
- Brinkhoff T., Kriegel H.-P., Schneider R., Seeger B.: *'Efficient Multi-Step Processing of Spatial Joins',* Proc. ACM SIGMOD Int. Conf. on Management of Data, Minneapolis, MN, 1994, pp. 197-208.
- Erwig M., Gueting R.H.: *"Explicit Graphs in a Functional Model for Spatial Databases"*, IEEE Transactions on Knowledge and Data Engineering, Vol.6, No.5, 1994, pp. 787-803.
- Ester M., Kriegel H.-P., Sander J.: *"Spatial Data Mining: A Database Approach"*, Proc. 5th Int. Symp. on Large Spatial Databases, Berlin, Germany, 1997, pp. 47-66.
- ESTER, M. et al., 2000, Spatial data mining: databases primitives, algorithms and efficient DBMS support. Data Mining and Knowledge Discovery, 4, 193-216
- Gueting R. H.: *"An Introduction to Spatial Database Systems"*, Special Issue on Spatial Database Systems of the VLDB Journal, Vol. 3, No. 4, October 1994.
- Guttman A.: *"R-trees: A Dynamic Index Structure for Spatial Searching",* Proc. ACM SIGMOD Int. Conf. on Management of Data, 1984, pp. 47-54.
- HAINING, R., 2003, Spatial Data Analysis: Theory and Practice (Cambridge: Cambridge University Press)
- Koperski K., Adhikary J., Han J.: *"Knowledge Discovery in Spatial Databases: Progress and Challenges"*, Proc. SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery, Technical Report 96-08, University of British Columbia, Vancouver, Canada, 1996.
- Koperski K., Han J.: *"Discovery of Spatial Association Rules in Geographic Information Databases"*, Proc. 4th Int. Symp. on Large Spatial Databases, Portland, ME, 1995, pp.47-66.
- Lu W., Han J.: *"Distance-Associated Join Indices for Spatial Range Search"*, Proc. 8th Int. Conf. on Data Engineering, Phoenix, Arizona, 1992, pp. 284-292
- Ng R. T., Han J.:*"Efficient and Effective Clustering Methods for Spatial Data Mining"*, Proc. 20th Int. Conf. on Very Large Data Bases, Santiago, Chile, 1994, pp. 144-155.
- Rotem D.: *"Spatial Join Indices"*, Proc. 7th Int. Conf. on Data Engineering, Kobe, Japan, 1991, pp. 500-509.