Categorization of Spatial Data Mining Techniques

Shashi Shekhar[*], Yan Huang[*], Jia-Wei Han[†], Sanjay Chawla[‡], Sucharita Gopal[§]

November 28, 2000

Abstract

Spatial data mining is the process of discovering interesting, previously unknown but potentially useful patterns from large spatial datasets. Extracting interesting and useful patterns from spatial datasets is more difficult than exatracting corresponding patterns from traditional numeric and categorical data due to the complexity of spatial data types and spatial relationships. A popular approach is to preprocessing the spatial components and apply classical data mining techniques. An alternative is to apply new models and objective functions. In this paper we provide a framework to understand the spatial data mining techniques in literature.

Keywords: spatial data mining,spatial databases, feature selection

[*] Computer Science Department, University of Minnesota 200 Union Street SE, Minneapolis, MN-55455 [*shekhar, huangyan*]@cs.umn.edu

[†] Simon Fraser University

[‡] Vignette Corp.

[§] Boston University

1 Introduction

Locations and their relations are important to understand many social, ecomomic and environmental phenomena. Furthermore, many scientific datasets in fields such as fluid dynamics, ecology and environmental are spatial. Spatial data mining is the process of discovering interesting, previously unknown but potentially useful patterns from large spatial datasets. The patterns found by spatial data mining can be used to fight crime, plan store locations, find customers, and protect environments. Data mining has been an active field for about a decade.Many interesting and useful patterns are found. Spatial datasets have multiple thematic maps with many objects in each layer, dozens of spatial relationships, and spatial auto-correlation properties. The question motivating this survey is how to treat spatial components of data effectively. In this survey, we focus on identifying the spatial aspects in the spatial data mining literature.

In section2, we review classical data mining patterns and then introduce spatial data mining patterns. Section 3.1 summarizes the spatial data mining literature focusing on feature selection, followed by a discussion of spatial concepts in models and objective functions in section 3.2 and section 3.3 respectively. Section 3.4 is another area which most researchers focus on, post processing. We'll have a short discussion in section 4 and summarize our conclusion in section 5.

2 Classical and Spatial Data Mining Patterns

Large databases are rich not only in data but in hidden information beyond human's ability to retrieve manually. Knowledge Discovery in Databases(KDD) has been defined as "The nontrivial extraction of implicit, previously unknown, and potentially useful information from data". Results in machine learning, statistics, and visualization have been extensively used to discover and present knowledge in a form which is comprehensible to humans.

The interest in data mining began in the early 1990s; since then data mining has identified several patterns. It will likely continue to be a fertile field with the interests in web mining, scientific mining, and finding more patterns increasing. Knowledge discovery from spatial databases started approximately at the same time. Complex data types and intrinsic relations between spatial components and non-spatial components as well as relationships between data themselves means that spatial data mining is more difficult but that it results in more interesting patterns.

In the following subsections, we summarize the patterns developed during the past several years in classical and spatial data mining along with spatial domain specific patterns.

2.1 Basic Concepts: Classical Data Mining Process and Patterns

Given raw data, Fig1 suggests an idealized sequence of the data mining process. Preprocessing and feature selection consist of operations that prepare data for subsequent analysis. This step includes compensating for the systematic errors of the raw data and isolating the most useful components for further inspection. The pattern selection step determines the type of pattern we wish to extract from the data. The result may be classification (supervised learning), clustering(unsupervised learning) or association rules. The different model alternatives and the objective functions of the selected patterns are considered for suitability and then the proper model and objective function are selected. Efforts are made to improve the efficiency and scalability of the selected algorithms. Finally, the dataset is plugged in and pattern is extracted. Postprocessing may be needed during or after pattern extraction to make patterns more resistent to noise/outliers. The extracted pattern is validated iteratively using known datasets. For some patterns, the previously described steps are sufficient while for others more these steps are just the beginning of knowledge discovery process. For example, for clustering analysis is needed to explain and take advantage of the resulting clusters.

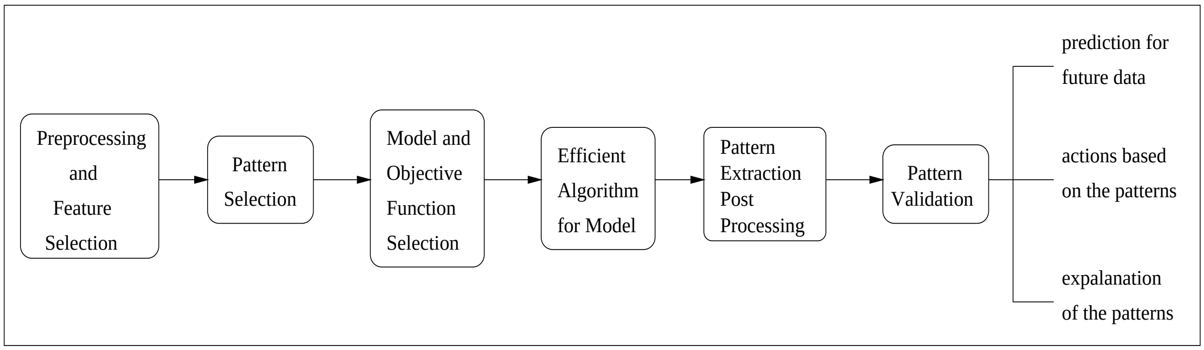

Figure 1: The Idealized Data Mining Process

The following subsections present the classical data mining patterns.

2.1.1 Classification

Given a set of data (training set) with one attribute as the dependant attribute, a model is built to predict the unknown dependent attribute of future data based on the other attributes as accurately as possible.

Classification usually uses the decision tree based approach(C4.5, ID3 etc.), the rule-set based approach, the bayesian approach, the instance based approach (k-Nearest Neighbor) or the neural network based approach.

A set of data with a dependent attribute known in advance (test set) is used to validate the model. Precision(P) of a particular class is defined as the number of correctly predicted instances of this class divided by the total number of instances predicted to be in this class. Recall(R) of a particular class is defined as the number of correctly predicted instances of this class divided by the total number of instances actually belonging to this class. To balance the extremes of models which favor one accuracy measure at the sot of decreasing the other, the F measure($F = 2 * P * R/(P + R)$) is introduced.

2.1.2 Clustering

Given a set of data, divide the data into k groups such that the data in one group are as similar as possible, and the data in different groups are as dissimilar as possible.

The partitioning based approach, e.g. K-Means, K-Mediod based, starts from K representitives of the K cluster(not necessarily actual points) and iteratively assigns each instance to one cluster in order to optimize an objective function. Hierarchical methods have two alternatives, namely divisive and agglomerative approaches. The divisive approach starts from an all-inclusive cluster and divide until only one-item clusters remain. The agglomerative approach starts from singleton clusters and combines the most similar clusters until only one all-inclusive cluster remains. Many other hybrid hierarchical clustering approaches exist.

2.1.3 Association Rules

Let $I = i_1, i_2, \ldots, i_m$ be a set of literals, called items. Let D be a set of transactions, where each transaction T is a set of items such that $T \subseteq I$. An association rule is of the form $X \Rightarrow Y, where X \subset I, Y \subset I, and XY = \phi. Pr(XY)/Pr(X)$ is called confidence of the rule and $Pr(XY)$ is called support of the rule [1].

The most famous example of an association rule is $Diapers \Rightarrow Beer$ which means "People buy dipers in the afternoon tend to buy some beer". The above is the basic association rule. Much research is looking at association rules with taxanomies, negative association rules, and indirect association rules.

2.1.4 Other Patterns

Apart from the above three most non-controversial labels of data mining patterns, other patterns such as sequential patterns, deviation/outlier detection etc. are being investigated too.

2.2 Interesting Properties of Spatial Data and Spatial Data Spatial Data Mining Patterns

Spatial data have traditional numeric and categorical data types as well as spatial data type. Spatial data types include points, line strings, polygons, and graphs. Spatial data are related to each other by three kind of intrinsic relations: topological, metric and directional [20]. These properties provide spatial data mining with more interesting patterns. All the above data mining patterns have their spatial counterparts. To illustrate, consider the fault map of the earth as an example . A fault divides the earth into two tectonic pieces. Geologic forces drive the locked pieces apart. When enough pressure builts, the jerky unlocking of the two pieces causes tragedy, e.g. Turkey's 1999 earthquake caused by the North Anatolian Fault. A fault's properties may include: type of fault, dimensions, start and end points, nearby communities, most recent rupture time, slip rate, and geometry. A geometry is a line string representing the fault. It is useful to group the faults which are similar together for future research focusing on different groups of faults [6, 7, 11, 18, 23]. If we overlay the fault map with other maps such as maps of mountains, or bodies of water, then association rules may help find interesting patterns such as [16]:

$$is\_a(X, normal\_fault) \wedge close\_to(X, sea) \rightarrow slip\_rate(X, high)(90\%)$$

The intrinsic relationships between spatial data also gives spatial data mining some unique patterns.

2.2.1 Spatial Characterization

Aggregate Proximity [12, 14]: Given a cluster of points, find efficiently K features (represented by polygons) that are closest to the majority of points in the cluster. Take the fault map as an example. Faults can be represented by line strings, but small faults can be represented by points (compared to water bodies, regions). Given a cluster of small faults, the aggregate proximity pattern finds the K polygons (representing water bodies, regions etc.) which are closest to most of the faults in the cluster.

The pattern is interesting because we can use it to explain the spatial properties of a given cluster.

[12, 14] also compare two clusters of points according to their common proximate features(represented by polygons).

Boundary Shape Matching [13]: Given a cluster of points, find K features (represented by polygons) which match the shape (totally or partially) of the cluster best.

[13] gives an example to demonstrate the this pattern: If part of a housing cluster follows the contour of a golf course, lake, or park, it makes sense to report this fact.

One might question the values of identifying the above distance and shape based patterns and argue that it is more effective to find these patterns by visualization and further more they are already known. But in other application domains such as ecology and bio-science, it is beyond human's ability to discover these patterns by visualization. Also, techniques are available to prune the uninteresting patterns [21].

2.2.2 Spatial Trend Detection

[5] introduces finding spatial trends using neighbor-hood graphs. [12] examines each object and paths connected to it within a range of length. [12] uses distance from the object as the independent variable and the difference of the non-spatial attribute as the dependent variable for the regression. There are two variations of the algorithms. Globle_Trend finds the trend which is applicable to all the paths leading away from the center object while Local_Trend finds the trend of each individual path. The spatial objects within a trend region, i.e., either the start objects or the objects along the trend paths, may be further investigated, for example, to find and explain the deviation (inverse trend) along some paths. In our fault map example, we may find that when leading away from the North Anatolian Fault, the slip rate decreases, or we may find that along some paths the slip rate actually increases. Such findings deserve our attention.

2.2.3 Other patterns

[4] characterizes a set of spatial data by finding the set of all tuples(non-spatial_attribute,value)for which the relative frequency in a given set, extended by its neighbor, is significantly different from the relative frequency in the database. For example, given a set of faults we may find these faults have a much greater slip rate than that of the whole faults database.

3 Spatial Aspects in Spatial Data Mining Literature

3.1 Spatial Feature Selection

Many attributes in the data may be irrelevant to the undergoing data mining task. For example, if we are interested in discovering patterns from a fault map, attributes such as type of fault, dimensions, start and end points, most recent rupture time, and slip rate may be most relevant while number of tours to the resort nearby may not be interesting to us. Furthermore, we may need to remove some redundant attributes. In machine learning this process is called feature selection.

The complexity of feature selection in spatial data mining occurs because spatial data are frequently related to each other. There are three basic types of spatial relationships: topological, metric and directional. These relationships are frequently handled by spatial data mining algorithms either implicitly or explicitly. Feature selection incorporates the spatial relationships implicitly by transforming spatial attributes into non-spatial attributes. There is a big set of spatial data mining problems in which the most difficult task is to transform spatial features into non-spatial features then plug them into classical knowledge discovery processes. Most of the spatial data mining work done today focuses on feature selection. The following paragraph provides a detailed summary of research done in spatial data mining focusing on spatial feature selection.

3.1.1 Feature Selection in Classification

[5] introduces neighbor-hood graph(path) concepts. Every object knows its neighbors based on the three basic relationships between spatial objects. When training the data, [5] adds some additional attributes to the data, namely the independent attributes of its neighbor. For example, when predicting the possibility of a fault's rupture in the next five years, the dataset attributes are extended by adding the neighbor's_slip_rate, neighbor's_dimension. Then the pre-precessed data can be fed into algorithms such as ID3 or C4.5. Essentially, while building the decision tree, [5]considers not only the attributes of the object to be classified but also the attributes of neighboring objects.

3.1.2 Feature Selection in Clustering

Most of the work done in spatial clustering focuses on spatial feature selection. Researchers either apply clustering algorithms to spatial attributes after spatial feature selection and collect and analyze the no-spatial components of the clusters, or they first classify the data based on non-spatial attributes and then apply clustering algorithm to each class.

[6, 11, 18, 23] introduce algorithms for spatial data mining while there is no strong evidence that they cannot be used in general clustering. [7] proposes some focusing techniques to speed up CLARANS [18]. CLARANS is a mediod-based randomized searching approach. One technique to reduce cost is to sample from R-tree and find clusters of the sampled data. The rest will be assigned to its nearest mediod. [7] shows that only the cluster represented by the mediod which is going to be replaced and cluster represented by its replacement mediod in next round are relevant. By partitioning the whole space into K parts using the K mediods(twice, once using the original mediods, once using the mediods after replacement), we can focus on the two relevant clusters. This kind of range query is effective by using R-tree.

3.1.3 Feature Selection in Association Rules

[16] introduces a method to mine all the association rules in the following form:

$$is\_a(X, big\_city) \wedge adjacent\_to(X, sea) \rightarrow close\_to(X, us\_boundary)(80\%)$$

where at least one of the predicates is a spatial predicate.

Two concept hierarchies are used: predicate hierarchy and data hierarchy. For example, predicate g_close_to is a generalization of spatial predicates: intersect, adjacent_to, contains and close_to. In the data concept hierarchy, Okanagan_lake may be generalized to larg_lake, lake, water along the path. Hierarchies are used for multi-level rules mining and fast processing.

Users are asked to specify the objects which they are interested in first in a form that specifies the main object of interest set and the relevant objects. An example is finding with the map of British Columbia the strong spatial association relationships between large towns and other "near_by" objects including mines, country boundary, water and major highways. The set of large towns is the main object set of interest while the mines, country boundary etc. are the relevant objects. [16] uses two-step computation: first generates association rule at a coarse level, e.g. , g_close_to, which is efficient by using R-tree or fast MBR techniques, and then only the objects with support higher than minimum support are passed to fine level(e.g. adjacent_to) rule generation. The association rules are derived straight from the high support itemsets. Concept hierarchies of data is used to generate multiple level association rules [8].

[16] maps each object in the main object set of interest to a transaction and spatial predicates to items. After feature selection, [16] uses Apriori [2] to find the rules.

3.2 Spatial Concepts in Models

[3] points out that current approaches towards solving spatial data mining problems are to use classical data mining tools after "materializing" spatial relationships and assuming identical independent distribution of data points. The fundamental property of spatial data is spatial autocorrelation, meaning the spatial data values are influenced by values in their immediate vicinity. The spatial autocorrelation model believes that spatially autocorrelated phenomena are the results of an underlying spatial structure. Ignoring this structure will result in poor models for spatial data. Take the fault map of the earth in 2.2 as an example. If we are given a fault map of the earth of the past and are asked to predict where the new faults will happen in the future, we may first impose a grid on the earth and then try to predict if there will be a fault in each cell. A fault represented by a line string may span across multiple cells. Dealing with spatial attributes other than points is not easy. For simplicity, fault spanning across cells is broken into separate smaller segments, each of which resides in a single cell . Since faults are result of many combinations of force, failure to the capture spatial autocorrelation will lead to model selection hat produces misleading results. Two alternative models are suggested in [3]:

- Spatial Error Model(SEM)

$$y = X\beta + \mu$$

$$\mu = \rho W\mu + \varepsilon$$

  where W is the contiguity matrix of the spatial configuration (use grid imposed) and ρ is a parameter that reflects the strength of spatial dependencies between the terms. In this model, spatial error terms are assumed to be spatially autocorrelated.

- Spatial Autoregressive Model(SAR)

$$y = \rho W y + X\beta + \varepsilon$$

  where we assume that the dependent values $y_{i'}$ are related to each other.

These two models cannot be applied directly because of the expensive computational cost and the non-trivial choice of the contiguity matrix W. More cooperative research with spatial statisticians is needed to explore this area. Problems remaining include finding more accurate spatial models, improving computation cost, modeling spatial components other than points (e.g., lines, polygons) and exploring spatial modeling in different spatial data mining patterns.

3.3 Spatial Terms in Objective Functions

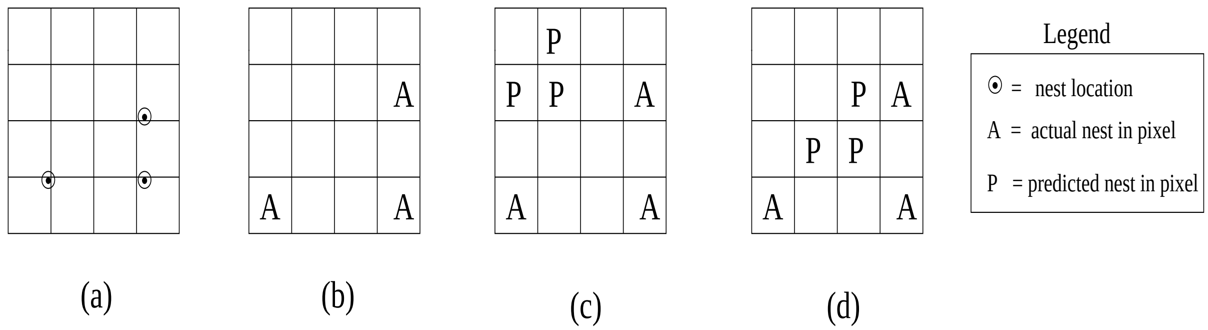

Figure 2: (a)The actual locations of faults, (b)Pixels with actual faults, (c)Location predicted by one model, (d)Location predicted by another model. Prediction(d) is spatially more accurate than (c). Classical measures of classification accuracy will not capture this distincition.

Classical accuracy measures such as precision, recall, F measure and Receiver Operating Characteristic (ROC) do not measure spatial accuracy. To illustrate, consider the example in 3.2 in Figure 2. Figure2[b] represents the actual faults after we impose a uniform grid and discretize continuous space into pixels. Classical accuracy measures cannot distinguish between the model which predicts Figure 2[c] and the model which predicts Figure 2[d]. But actually, prediction(d) is spatially more accurate than prediction(c) since the predicted fault locations are closer on average to some actual fault locations. Measures are needed to incorporate spatial accuracy. [3] suggests taking the average distance from every predicted fault location to its nearest actual location as a measure of spatial accuracy. Work in other spatial data mining patterns does not explicitly incorporate spatial accuracy. An interesting topic to explore is whether and how to measure spatial accuracy in other data mining patterns. For example, in the spatial association rule pattern, are the interest measures such as *Confidence, Life Factor, Gini Index and Piatetsky-Shapiro interest measure* sufficient to find spatially interesting rules and to prune spatially uninteresting rules?

3.4 Spatial Concepts in Algorithms

Domain specific techniques such as spatial access methods(e.g., R-tree), computational geometry techniques and spatial relationship handling methods can be used in spatial data mining algorithms to speed up computation or simply to find interesting patterns.

3.4.1 Algorithms Aimied at Improving Scalability

The SEM and SAR models [3] quantify spatial autocorrelation explicitly.The resulting models can be solved via complex numerical solvers or sampling based Markov Chain Monte Carlo method. The weaknesses of these methods are expensive computation and slow convergence. [3] uses a greedy search method to search parameter space. Starting from a random seed parameter-value tuple, [3] searches its immediate neighborhood parameter-value tuple in the given discretization and follows the neighbor which gives better outcome of objective function. The search continues until a local maximum is reached.

[7] makes data small enough to be memory resident to speed up the mediod based randomized search method CLARANS [18]by using the spatial access method R-tree. The first method is sampling. Because R-tree has a good property: minimize the overlap between directory rectangles, selecting the most central object from each R-tree data page will give us a good set of representatives. After clustering on the representatives, the remaining objects will be assigned to the nearest mediod. The second method is to focus on the relevant clusters. It has been shown that only the clusters represented by the current mediod which is going to be replaced in next round and the cluster represented by the replacement mediod are relevant. By finding the Voronoi polyhedron associated with these two mediods and using them as range to query the R-tree, the I/O cost is greatly reduced.

3.4.2 Algorithms Using Spatial Relationships

[5] argues that since spatial data mining algorithms use neighborhood relations extensively, we should build neighborhood graphs and neighborhood paths in advance and support some basic neighborhood graph(path) operations efficiently. A neighbor-hood graph $G_{neighbor}$ for some spatial relation "neighbor" is a graph (N,E) with the set of nodes N and set of edges E. Each node corresponds to an object of the database and two nodes $n_1 and n_2$are connected via some edge iff neighbor(object(n_1), object(n_2)). The predicate neighbor maybe be any of the basic relations, namely topological, metric and directional. [5] also defines several basic operations such as producing the neighborhood graph of an object regarding to a relation, getting all objects directly connected to an object satisfying some condition and argues that these operations are sufficient for KDD algorithms. To speed up the basic operations, neighbor-hood indexes are introduced. Neighbor-hood indexes can be the traditional B+ (R) tree with the leaves storing extra neighbor-hood object information.

3.4.3 Algorithms Using Computational Geometry

To find the Boundary Shape Matching Pattern, [13] approximates a cluster of points using an alpha-shape and then qualifies the matching degree of two facing curves by using the slope of the difference of each segment of the two curves.

To find the Aggregate Proximity Pattern, the algorithm CRH in [12, 14] uses a multiple filtering approach to find the top-k objects that are closest to the majority of the points in a cluster. To reduce the computational cost, the cluster shape and objects shapes are approximated by Encompassing Circles, Isothetic Rectangles, and Convex Halls in each filtering step.

3.4.4 Algorithms Using Concept Hierarchies

As in classical data mining, concept hierarchies are frequently used to reduce cost and find multiple-level rules (in association rules).

The predicate concept hierarchy is used in finding spatial association rules in [16] to allow the two step computation. Only the objects that passed the coarse level predicate and minimum support criteria will be investigated further. Data concept hierarchies are used in [16] to mine multiple-level rules.

To find the common nearest objects of n clusters of points based on the top-k nearest objects found for each of the n given clusters, [12, 14]uses a concept hierarchy to find the objects closed to most of the clusters. Concept hierarchies are used because it is unlikely that a specific object such as school is close to all the clusters but it is highly possible that each cluster is close to some private school, which is a generalization of school along the path in the concept hierarchy.

3.5 Spatial Concepts in Postprocessing

Postprocessing is another frequently used technique in spatial data mining. Techniques such as smoothing or ignoring minor outliers or island regions after merging the regions of some non-spatial characteristics are used. [19] use simulated annealing to separate composite clusters.

4 Discussion

The important properties of spatial data include spatial and non-spatial components, complex data types(e.g. points, lines, and polygons),intrinsic relationships between spatial data, multiple layer thematic map representations and spatial autocorrelation. The question for researchers in this area is which approach should we recommend, classical data mining with spatial feature selection or new spatial data mining approaches. As can be seen, most work in spatial data mining is focused on feature selection then apply the classical data mining techniques. Are feature selection, geometric computation, post processing and utilizing spatial access methods sufficient? Feature selection itself deserves much research effort. In the first approach, a set of typical feature selection methods needs to be identified. Techniques to handle spatial relationship needs to be created. In the second approach, effective methods to model spatial autocorrelation are needed as well as new spatial accuracy measures.

5 Conclusion

In this paper, we provide a framework to understand the spatial data mining literature. Two major approaches are identified: classical data mining method after feature selection and new spatial data mining approaches. As a conclusion, spatial data mining is an active and open field that is far from mature. Typical problems and corresponding solution frameworks still need to be defined and provided.

References

[1] R. Agrawal, T. Imielinski, and A. Swami. Mining association rules between sets of items in large databases. In *In Proc. of the ACM SIGMOD Conference on Management of Data, Washington, D.C.*, may 1993.

[2] R. Agrawal and R. Srikant. Fast algorithms for Mining associatin rules. In *VLDB*, may 1994.

[3] S Chawla, S. Shekhar, W.-L. Wu, and U. Ozesmi. modeling spatial dependencies for mining geospatial data. In *ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery 2000: 70-77*, 2000.

[4] M. Ester, A. Frommelt, H.P. Kriegel, and J. Sander. Algorithms for characterization and trend detection in spatial databases. In *In Proc. Fourth International Conference on Knowledge Discovery and Data Mining., AAAI Press, Menlo Park*, 1998.

[5] M. Ester, H.-P. Kriegel, and J. Sander. Spatial Data Mining: A Database Approach. In *In Proc. Fifth Symposium on Rules in Geographic Information Databases*, 1997.

[6] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *In Proc. Second International Conference on Knowledge Discovery and Data Mining, Portland, Oreg on. AAAI Press.*, 1996.

[7] M. Ester, H.-P. Kriegel, and X. Xu. Knowledge Discovery in Large Spatial Databases: Focusing Techniques for Efficient Class Identification . In *In Proc. Fourth International Symposium on Large Spatial Databases, Maine*, 1995.

[8] J. Han and Y. Fu. Discovery of multiple-level association rules from large databases. In *In Proc. 1995 Int. Conf. Very Large Data Bases, pages 420-431, Zurich, Switzerland*, September 1995.

[9] J. Han, M. Kamber, and A. Tung. Spatial clustering methods in data mining: a survey. In *do not know*, 1999.

[10] J. Han, K. Koperski, and N. Stefanovic. GeoMiner: A System Prototype for Spatial Data Mining. In *In Proc. ACM SIGMOD Conference on the Management of Data, Tucson, Arizona.*, 1997.

[11] I.-S. Kang, T.-W. Kim, and K.-J. Li. A Spatial Data Mining Method by Delaunay Triangulation. In *In Proc. Fifth ACM Workshop on Geographic Information Systems, Las Vegas, Nevada*, 1997.

[12] E.M. Knorr and R.T. Ng. Extraction of Spatial Proximity Patterns by Concept Generalization. In *In Proc. Second International Conference on Knowledge Discovery and Data Mining, Portland, Oregon. AAAI Press. 347-350*, 1996.

[13] E.M. Knorr, R.T. Ng, and D.L. Shilvock. Finding boundary shape matching relationships in spatial data. In *In Proc. 5th International Symposium, SSD 97. Springer-Verlag, Berlin, 29-46*, 1997.

[14] E.M. Knorr and Ng R.T. Finding aggregate proximity relationships and commonalities in spatial data mining. *IEEE Trans. Knowl. and Data Eng. 8(6):884-897*, 1996.

[15] K. Koperski, J. Adhikary, and J. Han. Knowledge Discovery in Spatial Databases: Progress and Challenges. In *In Proc. ACM SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery, Montreal, Canada. 55-70*, 1995.

[16] K. Koperski and J. Han. Discovery of Spatial Association Rules in Geographic Information Databases. In *In Proc. Fourth International Symposium on Large Spatial Databases, Maine. 47-66*, 1995.

[17] K Koperski and J. Han. An efficient two-step method for classification of spatial data. In *In Proc. Symposium on Spatial Data Handling(SDH'98), Vancouver, Canada, 1998*, 1998.

[18] R.T. Ng and J. Han. Efficient and effective clustering methods for spatial data mining. In *In Proc. Twentieth International Conference on Very Large Data Bases, Santiago, Chile*, 1994.

[19] J.-F. Roddick and M. Spiliopoulou. A Bibliography of Temporal, Spatial and Spatio-Temporal Data Mining Research. *SIGKDD Explorations 1(1): 34-38 (1999)*, 1999.

[20] S. Shekhar, S. Chawla, S. Ravada, A. Fetterer, X. Liu, and C.-T. Lu. Spatial Databases - Accomplishments and Research Needs. *TKDE 11(1): 45-55 (1999)*, 1999.

[21] A. Silberschatz and A. Tuzhilin. What Makes Patterns Interesting in Knowledge Discovery Systems. *TKDE*, 1999.

[22] E.-J. Son, I.-S. Kang, T.-W. Kim, and K.-J. Li. A spatial data mining method by clustering analysis. In *In Proc. Sixth International Symposium on Advances in Geographic Information Systems, GIS 98, 157-158*, 1998.

[23] T. Zhang, R. Ramakrishnan, and M. Livny. BIRCH: An Efficient Clustering Method for Very Large Databases. In *In Proc. ACM SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery, Montreal, Canada*, 1996.