Mining Spatial Data: Opportunities and Challenges of a Relational Approach

D. Malerba[1]

[1] Dipartimento di Informatica, Università degli Studi di Bari, Via Orabona 4, I-70126 Bari, Italy

Keywords: Spatial Data Mining, Multi-Relational Data Mining.

Abstract

Remote sensing and mobile devices nowadays collect a huge amount of spatial data which have to be analysed in order to discover interesting knowledge about economical, social and scientific problems. However, the presence of a spatial dimension adds some problems to data mining tasks. The geometrical representation and relative positioning of spatial objects implicitly define spatial relationships, whose efficient computation requires a tight integration of the data mining system with the spatial DBMS. The interactions between spatially close objects introduce different forms of autocorrelation whose effect should be considered to improve predictive accuracy of induced models and patterns. Units of analysis are typically composed of several spatial objects with different properties, and their structure cannot be easily accommodated by classical double entry tabular data. In the paper it is shown how these problems can be faced when a (multi-)relational data mining approach is considered for spatial data analysis. Moreover, the challenges that spatial data mining poses on current relational data mining methods are presented.

1 Introduction

In a large number of application domains, such as traffic and fleet management, environmental and ecological modeling, robotics, computer vision, and, more recently, computational biology and mobile computing, collected data present a spatial dimension. Indeed they are measurements on one or more attributes of objects which occupy specific locations. These (*spatial*) objects are characterized by a geometry (e.g., a line or an area) which is formulated by means of a reference system. This geometry implicitly defines both spatial properties, such as orientation, and spatial relationships of different nature, such as topological (e.g., intersects), distance or direction (e.g., north-of) relations. A *geographical* object represents a special case of spatial object whose relative position is specified with respect to the physical earth.

Studies in spatial data structures (Güting, 1994), spatial reasoning (Egenhofer and Franzosa, 1991), and computational geometry (Preparata and Shamos, 1985) have paved the way for the investigation of *spatial data mining*, which is related to the extraction of interesting and useful but implicit spatial patterns (Koperski et al., 1996). A spatial pattern expresses a spatial relationship among (spatial) objects and can take different forms such as classification rules, association rules, regression models, clusters and trends. Therefore, to extract spatial patterns from spatial data sets it is important to identify

1. the relevant spatial objects, and
2. the properties of, and relationships between, relevant spatial objects.

This makes the spatial data mining different from traditional data mining, where objects described in the data set are typically treated as independent observations.

2 What's special about spatial data mining

The main issues that characterize spatial data mining tasks are the following:

1. Spatial objects have a geometry which need to be represented. Although spatial phenomena are often inherently continuous, their digitization requires a representation in a discrete way by means of two types of data structures: *tessellation* and *vector* (Laurini and Thompson, 1992). The former partitions the space into a number of cells each of which is associated with a value of a given attribute. No variation is assumed within a cell and values correspond to some aggregate function (e.g., average) computed on original values in the cell. A grid of square cells is a special tessellation model called *raster*. This model is simple but the geometry of a spatial object is imprecise and requires large storage capabilities. In the vector model the geometry is represented by a vector of coordinates. This is a concise and precise representation but involved data structures are complex and the computation of spatial operations, such as intersection, is computationally demanding. Obviously, the formulation of a spatial data mining method cannot leave out of consideration the representation of the geometry.

2. Spatial objects have a locational property which implicitly defines spatial relationships between objects. The three main types of spatial relations are topological, distance and direction relations. Topological relations are invariant under homomorphisms, such as rotation, translation and scaling. Their semantics is precisely defined by means of the 9-intersection model proposed by Egenhofer and Franzosa (1991). Distance relations between points are typically computed on the basis of the Euclidean metrics, while the distance between two geometries (e.g., two areas) is defined by some aggregate function (e.g., the minimum distance between two points of the areas). Distance relations can be non-metric, especially when they are defined on the basis of a cost function which is not symmetric (e.g., the drive time). Directional relations can be expressed by the angle formed by two points with respect to the origin of the reference system, or by an extension of Allen's interval algebra which is based on projection lines (Mukerjee and Joe, 1990). The relational nature of spatial patterns makes the computation of these spatial relations crucial for the development of effective data analysis methods. To complicate matters, the user is generally interested in spatial patterns where relations are abstracted from the geometry of involved spatial objects (e.g., a river crosses a city, whatever their geometric representations are).

3. Spatial objects involved in spatial patterns often are of different type, and hence, have completely different sets of features which describe them. For instance, a town can be described in terms of economic and demographic factors, while a highway is described by the average speed limit, traffic and driving safety conditions, and so on. In spatial databases, objects of the same type are organized in *layers*, each of which can have its own set of attributes and at most one geometry attribute.

4. By picturing the spatial variation of some observed variables in a map, we may observe regions where the distribution of values is smoothly continuous with some boundaries possibly marked by sharp discontinuities. In this case, a variable is correlated with itself through space. *Spatial autocorrelation* is defined as the property of random variables taking values, at pairs of locations a certain distance apart, that are more similar (positive autocorrelation) or less similar (negative autocorrelation) than expected for randomly associated pairs of observations (Legendre, 1993). In geography, spatial autocorrelation is justified by Tobler's First law of geography, according to which "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970). However, spatial autocorrelation occurs in many other disparate fields, such as sociology (e.g., social relations affect social influence), web mining (e.g., hyperlinked web pages typically share the same topic), and bioinformatics (e.g., proteins located in the same place in a cell are more likely to share the same function than randomly selected proteins). In statistics spatial autocorrelation is distinguished in two primary types: *spatial error* (correlations across space in the error term), and *spatial lag* (the dependent variable in space 'i' is affected by the independent variables in space 'i' as

well as those, dependent or independent, in space 'j'). Most statistical models are based on the assumption that the values of observations in each sample are independent of one another, but spatial autocorrelation (or *spatial dependence*, as typically called in statistics) clearly indicates a violation of this assumption. As observed by LeSage and Pace (2001), "anyone seriously interested in prediction when the sample data exhibit spatial dependence should consider a spatial model", since this can take into account different forms of spatial autocorrelation. In addition to predictive data mining tasks, this consideration can also be applied to descriptive tasks, such as spatial clustering or spatial association rule discovery. More in general, the analysis of spatial autocorrelation is crucial and it can be fundamental to build a spatial component into (statistical) models for spatial data. The inappropriate treatment of sample data with spatial dependence could obfuscate important insights and observed patterns may even be inverted when spatial autocorrelation is ignored (Kühn, 2007).

Traditional data mining algorithms offer inadequate solutions to all these issues. They do not deal with spatial data characterized by a geometry, do not handle observations of different type, do not naturally represent spatial relationships between observations, nor take them into account when mining patterns.

To overcome some of these limitations, several extensions have been investigated in spatial statistics, where spatial dependence is typically modeled by the following linear models (LeSage and Pace, 2001):

$$y = X\alpha + \beta Dy + DX\gamma + \epsilon$$

where y is the $n \times 1$ vector of observations of the dependent (or response) variable, α considers the influence of the independent (or explanatory) variables observed in 'i' on the response variable in 'i', D is a spatial weight matrix with elements $D_{ij} > 0$ for observations 'j' sufficiently close (as measured by some metric) to observation 'i' (D defines the neighborhood), β reflects the strength of the spatial dependence on the response variable of the neighbors, γ reflects the strength of the spatial dependence on the explanatory variables of the neighbors, and ϵ reflects "noise" or a stochastic disturbance in the spatial dependence relation.

However, the application of these spatial models still present some problems. First, the neighborhood matrix D has to be carefully defined in order to specify to what extent a spatially close observation in space 'j' can affect the response observed in 'i'. With a proper choice of D, the residual error should, at least theoretically, have no systematic variation. Second, it is unclear how D can express the contribution of different spatial relationships, such as a polluting industry in an "adjacent" area and a highway "crossing" the same area. Third, spatial dependencies are all handled in a pre-processing or feature extraction step which typically ignores the subsequent data mining step. In principle, a data mining method which can handle spatial dependencies directly, presents the advantage of considering only those dependencies that are really relevant for the task at hand. Fourth, all spatial objects involved in a spatial phenomena (rows of the matrix X) are uniformly represented by the same set of attributes. This can be a problem when spatial objects are of different types and are characterized by different properties. Fifth, there is no clear distinction between the *reference* (or target) *objects*, which are the main subject of analysis, and the *task-relevant objects*, which are spatial objects "in the neighborhood" that can contribute to explain the spatial variation.

3 Opportunities for a Relational Approach

Problems reported above are due to the fact that in spatial data mining the units of analysis are typically composed of several spatial objects with different properties, and that their structure cannot be easily accommodated into a classical double-entry table whose columns correspond to elementary (nominal, ordinal or numeric) single-valued attributes. In fact, spatial data sets can be naturally modeled as a set of relations $R_1, \ldots, R_n$, such that each relation R_i has a number of elementary attributes $A_1^i, \ldots A_{m_i}^i$ and possibly a geometry attribute G_i (relations with geometry attributes are the *layers*). Joins between relations R_i are either explicitly modeled by means of foreign key constraints or implicitly represented by spatial joins $R_i \bowtie_\theta R_j$, where R_i and R_j are two layers, and θ is a binary predicate, such as intersects, contains, northwest, adjacent, to cite some, which is evaluated on the geometry attributes G_i and G_j (Shashi and Chawla, 2003).

Example. To investigate the social effects of public transportation in a British city, a spatial data set made of three relations is considered. The first relation *ED* contains information on enumeration districts, which are the smallest areal units for which census data are published in UK. In particular *ED* has two attributes, the identifier of an enumeration district, and a geometry attribute (a closed polyline) which describes the area covered by the enumeration district. The second relation *BL* describes all the bus lines which cross the city. In this case, relevant attributes are the name of a bus line, the geometry attribute (a line) representing the route of a bus and the type of bus line (classified as main or secondary). The third relation *CE* contains some census data relevant for the problem, namely, the number of households with 0, 1, or '> 1' cars. This relation also includes the identifier of the enumeration district, which is a foreign key for the table *ED*. A unit of analysis corresponds to an enumeration district (the reference object), which is described in terms of number of cars per household and crossing bus lines (bus lines are the task-relevant objects). The relationship between reference objects and task-relevant objects is established by means of a spatial join which computes the intersection between the two layers *ED* and *BL*. This relationship allows us to discover truly *relational patterns*, such as "the enumeration districts with a high percentage of households which own less than two cars, are served by at least two bus lines, one of which is a main bus line." Here the verb 'served' is purposely introduced, to show that spatial patterns of interest may not necessarily be expressed in terms of the original spatial predicates used in the spatial join operations. The most obvious interpretation of this verb can be the topological relation 'intersect' between the area of an enumeration district and the bus line, although other more sophisticated interpretations are possible (e.g., on the basis of the length of the intersected line). However, it may well be the case that an enumeration district with few households owning less than two cars is not actually crossed by a bus line, but rather it is spatially surrounded by several other enumeration districts where all conditions above hold. In this case, to take this spatial autocorrelation into account, a spatial join between *ED* and itself can be computed and the relational patterns can be searched *across* the units of analysis.

Previous example shows that *(multi-)relational data mining* (MRDM) offers the most suitable setting for spatial data mining tasks. Indeed, MRDM tools can be applied directly to data distributed on several relations to find relational patterns which involve multiple relations (Džeroski and Lavrač, 2001). Relational patterns can be expressed in SQL, but also in first-order logic (or predicate calculus), which explains why many MRDM algorithms originate from the field of inductive logic programming (ILP) (Muggleton, 1992; De Raedt, 1992; Lavrač and Džeroski, 1994).

Upgrading a classical data mining algorithm devised for double-entry tabular data to a relational setting is not a trivial task (Van Laer and De Raedt, 2001). For instance, it may be necessary to extend the definition of distance measure to data distributed among several tables. Efficiency is also very important, as even testing a given relational pattern for validity is often computationally expensive. Moreover, for relational pattern languages, the number of possible patterns can be very large and it is necessary to constrain the search space by means of some form of "declarative bias". An exhaustive list of theoretical results and techniques that have been developed to improve the efficiency and scalability of MRDM approaches is reported in (Blockeel and Sebag, 2003).

4 Challenges for a Relational Approach

Although the MRDM setting seems the most suitable for spatial data mining, there are still several challenges that must be overcome and issues that must be resolved before the relational approach can be effectively applied to spatial data mining. Some of them are reported in the following:

1. Many MRDM methods do take advantage of knowledge on data model (e.g., foreign keys), which is obtained free of charge from the database schema, in order to guide the search process. However, this approach does not fit well spatial databases, since the database navigation is also based on the spatial relationships which are not explicitly modeled in the schema. To

solve this problem spatial relationships can be computed and explicitly represented during the pre-processing step of the knowledge discovery process. This approach is typically followed by statisticians before computing the spatial weight matrix D. It has also been adopted in the GeoMiner system (Han et al., 1997) whose data mining algorithms, though, operate on a single database relation obtained from the preprocessing step. Also Ester et al. (1999) propose to precompute distance, direction and topological relations and to materialize (i.e., store) them into some database relations (called neighborhood indices), which are then used by data mining algorithms to efficiently retrieve all neighbors (with respect to some spatial relation) of a given spatial object. A feature extraction module is implemented into the ARES system (Appice et al., 2005) to precompute spatial relationships which are converted into Prolog facts used by the ILP system SPADA (Malerba and Lisi, 2001) to generate spatial association rules. The precomputation is justified by the fact that spatial databases are rather static, since there are not many updates on objects such as geographic maps. However, the number of spatial relationships between two layers can be very large and many of them might be unnecessarily extracted. The alternative is to dynamically perform spatial joins only for the part of the hypothesis space that is really explored during search by a data mining algorithm. This approach has been implemented in two MRDM systems, namely SubgroupMiner for subgroup mining (Kloesgen and May, 2002) and Mrs-SMOTI for regression analysis (Malerba et al., 2005). Both systems realize a tight integration with a spatial DBMS (namely, Oracle Spatial) and have been applied to datasets where few spatial relationships are actually computed. Spatial index structures, such as R-trees (Guttman, 1984), are used to speed up the processing of spatial joins. However, scalability remains a problem when many spatial predicates have to be computed. A scalability issue arises also in spatial statistics, since the spatial weight matrix D can be very large and sparse (LeSage and Pace, 2001).

2. Although the presence of autocorrelation in spatial phenomena strongly motivates a MRDM approach to spatial data mining, it also introduces additional challenges. It particular it has been proven that the combined effect of autocorrelation and concentrated linkage (i.e., high concentration of objects linked to a common neighbor) can bias feature selection in relational classification (Jensen and Neville, 2002). In particular, the distribution of scores for features formed from related objects with concentrated linkage presents a surprisingly large variance when the class attribute has a high autocorrelation. This large variance causes feature selection algorithms to be biased in favor of these features, even when they are not related to the class attribute, that is, they are randomly generated. Conventional hypothesis tests, such as the χ^2-test for independence, which evaluate statistically significant differences between proportions for two or more groups in a data set, fail to discard uninformative features. Indeed, they are based on the i.i.d. assumption, while observations drawn from a relational data set may not be independent. Most of MRDM algorithms do not account for this bias, a notable exception being a relational probability tree learning algorithm that uses a randomization test to adjust for feature selection bias (Neville et al., 2003). Pseudosamples are generated from the relational data by retaining the linkage present in the original sample and the autocorrelation among the class labels, and, at the same time, by destroying the correlation between the original attributes and the class labels. Therefore, pseudosamples are appropriately conform to the null hypothesis and can be used to estimate a p-value for the actual data.

3. Inductive learning algorithms designed for predictive tasks may require large sets of labeled data. However, the common situation is that only few labeled training data are available for mining, although a very large test set must be classified. This is especially true in geographical data mining, where large amounts of unlabeled geographical objects (e.g., map cells) are available and manual annotation is fairly expensive. Inductive learning algorithms would actually use only the few labeled examples to build a prediction model, thus discarding a large amount of information potentially conveyed by the unlabeled instances. The idea of *transductive inference* (or transduction) (Vapnik, 1998) is to analyze both the labeled (training) data and the unlabeled (working) data to build a classifier and classify (only)

the unlabeled data as accurately as possible. Transduction is based on a (semi-supervised) smoothness assumption according to which if two points x_1 and x_2 in a high-density region are close, then so should be the corresponding outputs y_1 and y_2 (Chapelle et al., 2006). However, in spatial domains where closeness of points corresponds to some spatial distance measure, this assumption is implied by (positive) spatial autocorrelation. Therefore, the transductive setting seems especially suitable for spatial classification and regression, and more in general, for those relational learning problems characterized by autocorrelation on the dependent variables. Only recently, a work on the transductive relational learning has been reported in the literature (Ceci et al., 2007), and some preliminary results on spatial classification tasks show the effectiveness of the transductive approach (Appice et al., 2007).

4. In predictive data mining tasks the generation of patterns which express the spatial autocorrelation of the dependent variable raises the issue of how inference on new cases should be performed. Indeed, these patterns take the form:

   $y_i = f(x_i, x_{N(i)}, y_{N(i)})$

   where y_i (x_i) is the value of the dependent (independent) variable in space 'i' while $y_{N(i)}$ ($x_{N(i)}$) represents the value(s) of the dependent (independent) variable for the i's the neighbor(s). For instance, the price level for a good at a retail outlet in a city depends on the price for the same good in the nearby. In order to predict y_i it is necessary to know the value(s) of $y_{N(i)}$, which might be unavailable (the related values of the dependence variable are to be inferred as well). In this case both y_i and all unknown values $y_{N(i)}$ have to be inferred *collectively.* A possible approach to collective inference combines locally-learned individual inference models with a joint inference procedure (e.g., relaxation labeling) to make an inference. An example is represented by iterative classification (Neville and Jensen, 2000), which dynamically updates the attributes of some objects as inferences are made about related objects. Inferences made with high confidence in initial iterations are fed back into the data and are used to inform subsequent inferences about related objects. Iterative classification works well when the classification model allows us to make initial inferences accurately, otherwise all subsequent predictions will be misled due to a ripple effect. An alternative approach is given by *joint relational models*, which first estimate the joint probability distribution over the variables of objects both in i and in $N(i)$ and then jointly infer the values of both y_i and $y_{N(i)}$. In particular, probabilistic relational models can be used to represent a joint probability distribution over the attributes of a relational dataset (Getoor et al., 2001, Neville and Jensen, 2003). By making inferences about multiple instances simultaneously, joint inference can exploit autocorrelation in the data to improve predictions (Jensen et al., 2004). Therefore, this inference procedure seems particularly suitable for spatial data sets and should be better investigated in the context of spatial data mining.

5. Spatial objects are often organized in hierarchies. By descending/ascending through a hierarchy it is possible to view the same spatial object at different levels of abstraction (or granularity). Spatial patterns involving the most abstract spatial objects can be well supported but at the same time they are the less confident. Therefore, spatial data mining methods should be able to explore the search space at different granularity levels in order to find the most interesting patterns (e.g., the most supported and confident). In the case of granularity levels defined by a containment relationship, this corresponds to exploring both global and local aspects of the underlying phenomenon. Very few data mining techniques do automatically support this multiple-level analysis. In general, the user is forced to repeat independent experiments on different representations, and results obtained for high granularity levels are not used to control search at low granularity levels (or viceversa). Two noticeable exceptions are represented by Geo-associator (Koperski and Han, 1995), a module of GeoMiner which mines spatial association rules from data represented in a single relation (table) of a relational database, and SPADA (Malerba and Lisi, 2001), which discovers multi-level spatial association rules from relational data. SPADA has also been used in an associative classification framework: once strong spatial association rules with only the class

label in the consequent are extracted for each granularity level, they are used to mine either propositional or structural spatial classifiers (Ceci et al., 2004; Ceci and Appice, 2006).

6. A large amount of knowledge is often available on spatial phenomena. This is particularly true in the special case of geographic knowledge discovery, where relations among spatial objects express natural geographic dependencies (e.g., a port is adjacent to a water body). These dependences are expressed in non-novel or uninteresting patterns but with a very high support and confidence. If this geographic knowledge were used to constrain the search for new patterns, the scalability of the spatial data mining algorithms would greatly increase. Actually, these dependencies are represented either in geographic database schemas through one-to-one and one-to-many cardinality constraints or in geographic ontologies. Therefore, their usage can be done at no additional cost in a multi-relational data mining perspective, thus moving a step-forward toward knowledge-rich data mining (Domingos, 2007). In the context of spatial data mining, both Appice et al. (2005) and Bogorny et al. (2006) explain how to use knowledge to constrain the search space for spatial association rules.

7. Spatial reasoning is the process by which information about objects in space and their relationships are gathered through measurement, observation or inference, and used to arrive to valid conclusions regarding the objects relationships. For instance, in spatial reasoning, the accessibility of a site A from a site B can be recursively defined on the basis of the spatial relationships of adjacency or contiguity. Principles of spatial reasoning have been proposed for both quantitative and qualitative approaches to spatial knowledge representation. Quantitative spatial reasoning deals with exact numerical values, such as coordinates and distances, and are more akin to machine reasoning, while qualitative spatial reasoning (Freksa, 1991) deals with abstract representations (e.g., 'northwest' and 'far') and is more closely related to the way humans reason. Qualitative spatial reasoning is arguably efficient and can also deal to some extent with imprecision, uncertainty, and incompleteness, which quantitative reasoning cannot. Embedding spatial reasoning in spatial data mining is crucial to make the right inferences either when patterns are generated or when patterns are evaluated. Surprisingly, there are few examples of data mining systems which support some form of spatial reasoning. In SPADA, a limited form of spatial inference is supported if rules of spatial reasoning are encoded in the background knowledge (Malerba et al., 2002). In particular SPADA applies an ILP technique, known as "saturation", to make explicit those pieces of information that are implicit in the spatial units of analysis, given the background knowledge. However, although a general-purpose theorem prover for predicate logic can be used for spatial reasoning (as in SPADA), constraints which characterize the spatial problem solving have to be explicitly formulated in order to make the semantics consistent with the target domain 'space'. Therefore, embedding specialized spatial inference engines in the spatial data mining systems seems to be the most promising, but still unexplored, solution.

Obviously, this list of challenges is not exhaustive, but rather it is indicative of the possible synergies between two research fields, namely spatial data analysis and multi-relational data mining, which have been developed independently in the communities of statisticians and computer scientists, but which share surprisingly many research issues. It is hard to envisage whether the multi-relational approach represents the next step in spatial data mining. The author's hope is to contribute bridging the understanding between the fields by providing some references to work already done by the different communities. Que sera, sera.

References

Appice, A., Berardi, M., Ceci, M. and Malerba, D. (2005). Mining and filtering multi-level spatial association rules with ARES. In M.S. Hacid, Z.W. Ras and S. Tsumoto (Eds.), *15th International Symposium On Methodologies For Intelligent Systems, ISMIS 2005*, LNAI 3488, Springer-Verlag, 342–353.

Appice, A., Barile, N., Ceci, M., Malerba, D. and Singh, R.P. (2007). Mining geospatial data in a transductive setting. In: Zanasi, A., Brebbia, C.A., Ebecken, N.F.F. (Eds.), *Data Mining VIII*, WIT Press, Southampton, 141–150.

Blokheel, H. and Sebag, M. (2003). Scalability and efficiency in multi-relational data mining. *SIGKDD Explorations*, **5**(1) 17-30.

Bogorny, V., Camargo, S., Engel, P. and Alvares L.O. (2006). Mining frequent geographic patterns with knowledge constraints. *Proc. of the 14th annual ACM Int. Symposium on Advances in Geographic Information Systems*, 139–146.

Ceci, M., Appice, A. and Malerba, D. (2004) Spatial associative classification at different levels of granularity: A probabilistic approach. In J.-F. Boulicaut, F. Esposito, F. Giannotti, and D. Pedreschi (Eds.), *Knowledge Discovery in Databases: PKDD 2004*, LNAI, 3202, Springer-Verlag, 99-111.

Ceci, M. and Appice, A. (2006). Spatial associative classification: Propositional vs. structural approach. *Journal of Intelligent Information Systems*, **27**(3) 191–213.

Ceci, M., Appice, A., Barile, N. and Malerba, D. (2007). Transductive Learning from Relational Data. In: P. Perner (Ed.), *Machine Learning and Data Mining in Pattern Recognition*, LNAI 4571, Springer, Berlin, 324-338.

Chapelle O., Schölkopf B. and Zien A. (Eds.) (2006). *Semi-supervised learning*, MIT Press, Cambridge:MA.

De Raedt, L. (1992). *Interactive theory revision*, Academic Press, London.

Domingos, P. (2007). Toward knowledge-rich data mining. *Data Mining and Knowledge Discovery*, **15**(1) 21–28.

Džeroski, S. and Lavrač, N. (2001). *Relational data mining*, Springer-Verlag, Berlin.

Egenhofer, M.J. and Franzosa, R. (1991). Point-set topological spatial relations. *International Journal of Geographical Information Systems*,**5**(2) 161-174.

Ester, M., Kriegel, H.-P., and Sander, J. (1999). Knowledge discovery in spatial databases. *KI-99: Advances in Artificial Intelligence*, LNCS 1701, Springer, Berlin, 61–74.

Freksa, C. (1991) Qualitative spatial reasoning, *Proceedings of the Conference on Cognitive and Linguistic Aspects of Geographic Space*, 361–372.

Getoor, L., Friedman N., Koller D. and Pfeffer, A. (2001) Learning probabilistic relational models. In: S. Džeroski and N. Lavrač (Eds.), *Relational Data Mining*, Springer-Verlag, Berlin, 307–335.

Güting, R.H. (1994). An introduction to spatial database systems, *VLDB Journal*,**4**(3) 357-399.

Guttmann, A. (1984). R-trees: A dynamic index structure for spatial searching. *Prof. of the ACM SIGMOD Int. Conf. on Management of Data*, 47–57.

Han, J., Koperski, K. and Stefanovic, N. (1997). GeoMiner: A System Prototype for Spatial Data Mining. *Prof. of the ACM SIGMOD Int. Conf. on Management of Data*, 553–556.

Jensen, D., and Neville, J. (2002). Linkage and autocorrelation cause feature selection bias in relational learning. *Proc. of the Nineteenth Int. Conf. on Machine Learning*, 259–266.

Jensen, D., Neville, J., and Gallagher, B. (2004). Why collective inference improves relational classification. *Proc. of the 10th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, 593-598.

Klösgen, W., and May, M. (2002). Spatial subgroup mining integrated in an object-relational spatial database. In: T. Elomaa, H. Mannila and H. Toivonen (Eds.), *Principles of Data Mining and Knowledge Discovery (PKDD), 6th European Conference*, LNAI 2431, Springer-Verlag, Berlin, 275-286.

Koperski, K. and J. Han (1995). Discovery of spatial association rules in geographic information databases. In *Advances in Spatial Databases*, LNCS 951, Springer-Verlag, 47–66.

Koperski, K., Adhikary, J. and Han, J. (1996). Spatial data mining: Progress and challenges, *Proc. ACM SIGMOD Workshop on Research Issues on Mining and Knowledge Discovery*, 1–10.

Kühn, I (2007). Incorporating spatial autocorrelation may invert observed patterns. *Diversity and Distributions, (Diversity Distrib.)*, **13** 66-69.

Laurini, R. and Thompson, D. (1992). *Fundamentals of spatial information systems.* Number 37 in APIC Series. Academic Press, London.

Lavrač, N. and Džeroski, S. (1994). *Inductive logic programming: techniques and applications*, Ellis Horwood, Chichester.

Legendre, P. (1993). Spatial autocorrelation - Trouble or new paradigm. *Ecology*, **74**, 1659-1673.

LeSage J.P., and Pace, K. (2001). Spatial dependence in data mining. In: R.L. Grossman, C. Kamath, P. Kegelmeyer, V. Kumar, and R.R. Namburu (Eds.), *Data Mining for Scientific and Engineering Applications*, Kluwer Academic Publishing, Norwell:MA, 439–460.

Malerba, D., and Lisi, F.A. (2001). Discovering associations between spatial objects: An ILP application. In: C. Rouveirol and M. Sebag, *Inductive Logic Programming, 11th International Conference, ILP 2001*, LNCS 2157, Springer, Berlin, 156–163.

Malerba, D., Esposito, F., Lisi, F.A., and Appice, A. (2002). Mining Spatial Association Rules in Census Data. *Research in Official Statistics*, **5**(1) 19–44.

Malerba, D., Ceci, M., and Appice A. (2005). Mining Model Trees from Spatial Data. In: A. Jorge, L. Torgo, P. Brazdil, R. Camacho and J. Gama (Eds.), *European Conference on Principles and Practice of Knowledge Discovery in Databases*, LNAI 3721, Springer, Berlin, 169–180.

Muggleton, S. (1992). *Inductive logic programming*, Academic Press, London.

Mukerjee, A. and Joe, G. (1990). A qualitative model for space. *Proceedings of AAAI-90*,AAAI Press, 721–727.

Neville, J., and Jensen, D. (2000). Iterative classification in relational data. *Proc. of the AAAI-00 Workshop on Statistical Relational Learning*, 42-49.

Neville, J. and Jensen, D. (2003). Collective classification with relational dependency networks. *Proceedings of KDD-2003 Workshop on Multi-Relational Data Mining*, 77–91.

Neville, J., Jensen, D., Friedland, L. and Hay M. (2003). Learning relational probability trees. *Proc. of Ninth ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, 625–630.

Preparata, F. and Shamos, M. (1985). *Computational geometry: An introduction.* Springer-Verlag, New-York.

Shekhar, S. and Chawla, S. (2003). *Spatial databases: A tour.* Prentice Hall, Upper Saddle River:NJ.

Tobler, W. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, **46** 234-240.

Van Laer, W. and De Raedt, L. (2001) How to upgrade propositional learners to first order logic: a case study. In: S. Džeroski and N. Lavrač (Eds.), *Relational Data Mining*, Springer-Verlag, Berlin, 235–261.

Vapnik, V. (1998). *Statistical Learning Theory.* Wiley, New York.