# Mining Spatial Association Rules in Census Data

Donato MALERBA, Floriana ESPOSITO and Francesca A. LISI
*Dipartimento di Informatica, University of Bari*
*Via Orabona 4 – 70126 Bari, Italy*
*e-mail:malerba@di.uniba.it*

**Abstract:** This paper is related to part of the work done within the IST project SPIN! (Spatial Mining for Data of Public Interest) managed by Eurostat. The goal is that of extending methods and techniques based on the theory of computational logic to the discovery of association rules in geo-referenced statistical data. A novel technique is presented and preliminary results on its application to census data from Stockport, one of the ten Metropolitan Districts of Greater Manchester, UK, are reported.

**Keywords:** Data mining, geo-referenced statistical data, association rules.

## 1. Introduction

In some works on spatial representation from the social scientist's perspective [17], socioeconomic phenomena have been conceptualized as *spatial objects*, without assuming any particular application such as marketing or resource allocation. Spatial objects in this sense are entities having both spatial location and spatially independent attribute characteristics [8]. Population data are among the potentially spatial socioeconomic data. They are most commonly related to geographic locations by reference to areal spatial objects such as census zones, electoral constituencies, local government areas, or regular grid squares. In the UK, for instance, the smallest areal unit for which census data are published is nowadays the enumeration district (ED). The digital ED boundaries produced for the 1991 UK census enable the spatial representation of census data in the computer databases. Indeed, population censuses of the 1990s provided an added impetus to the application of geographical information systems (GIS) to socioeconomic uses. These advances in the practice of geo-referencing census data also caused a growing demand for more powerful data analysis techniques that can analyze population data as spatial objects.

Statistical spatial analysis has been the most common approach to the analysis of geo-referenced data [18, 20]. Being a well-studied area, it supplies a large number of algorithms including various optimization techniques. It handles numerical data very well and usually comes up with realistic models of spatial phenomena. The statistical approach to spatial analysis shows some weaknesses in dealing with dependency in spatial data. Most methods are exploratory and when applied to spatially correlated data some of them are of unknown reliability having been developed initially, like so many areas in statistics, for situations where observations are independent [9]. This contrasts with the nature of spatial data where spatial objects are influenced by their neighboring objects as pointed out by [7]. In recent times, alternative approaches to spatial analysis have been emerged. In particular, the extension of data mining methods and techniques to spatial databases has been attempted to allow *the extraction of implicit knowledge, spatial relations, or other patterns not explicitly stored in spatial databases* [11]. Extracted knowledge can take on various forms according to the spatial data mining task at hand (discrimination, characterisation, clustering, classification, etc.). We are concerned with the task of mining *spatial association rules*, namely the detection of associations between spatial objects [12].

In this paper we propose the application of logic-based methods and techniques to the discovery of spatial association rules. In particular, we resort to Inductive Logic Programming (ILP) which is a form of inductive learning derived from the theory of computational logic [13]. Computational logic relies on an augmented expressive power which allows us to represent spatial relations and symbolic background knowledge (such as spatial hierarchies, spatial constraints and rules for spatial qualitative reasoning) in a very elegant and natural way. Thus, it enables applications which can not be tackled by traditional statistical techniques in spatial data analysis. The technique being proposed has been implemented in the ILP system SPADA (Spatial Pattern Discovery Algorithm) [15].

This work is in partial fulfillment of the research objectives set by the IST project SPIN! (Spatial Mining for Data of Public Interest) funded by the European Union (`http://www.ccg.leeds.ac.uk/spin/`). From the application side, the aim of the project is the development of an integrated interactive Internet-enabled spatial data mining system to enhance decision-making based on spatio-temporally referenced data and publish geographical data mining services on the WWW. The application to census data from Stockport - one of the ten Metropolitan Districts of Greater Manchester, UK - is the first step towards the fulfillment of the main social objective of the SPIN! Project, namely to put to practical use the timely, cost-effective dissemination of statistical information over the Internet.

The paper is organized as follows. Section 2 will introduce the task of mining spatial association rules, by discussing some issues raised by the application of computational logic to spatial data mining and the solutions adopted by the system SPADA. In Section 3 we report preliminary results on the application of SPADA to Manchester Stockport census data. Conclusions and future work will be drawn in Section 4.

## 2. Mining spatial association rules with SPADA

The discovery of spatial association rules is a descriptive mining task aiming at the detection of associations between *reference objects* and some *task-relevant objects*, the former being the main subject of the description while the latter being spatial objects that are relevant for the task at hand and spatially related to the former. For instance, we may be interested in describing a given area by finding associations among large towns (reference objects) and spatial objects in the road network, hydrography and administrative boundaries layers (task-relevant objects). The task usually involves massive spatial computation to extract spatial relations from the underlying spatial database. Some kind of taxonomic knowledge on task-relevant geographic layers may also be taken into account to get descriptions at different concept levels (*multiple-level association rules*). As usual in the problem setting of association rule mining, we search for associations with large support and high confidence (*strong rules*).

The problem of mining spatial association rules can be formally stated as follows:

*Given*

- a spatial database (SDB),
- a set of reference objects *S*,
- some task-relevant geographic layers $R_k$, $1 \leq k \leq m$, together with spatial hierarchies defined on them,
- a couple of thresholds for each level *l* in the spatial hierarchies, *minsup*[*l*] and *minconf*[*l*]

*Find* strong multiple-level spatial association rules.

The mining task at hand has been already tackled by Koperski and Han [12]. They propose a top-down, progressive refinement method which exploits taxonomies both on topological relations and spatial objects. The method has been implemented in the module Geo-associator of the spatial data mining system GeoMiner [10].

We propose an upgrade of Koperski's method to more powerful representations of data. The proposal is inspired to the work on multi-relational data mining done by Dehaspe and De Raedt [3]. The basic idea is that a spatial database boils down to a set of logic-based facts once the spatial relationships between reference objects and task-relevant objects have been extracted. In particular, we adopt Datalog [2] as a representation formalism. Datalog is a logic language for relational databases which allows us to represent spatial relations and symbolic background knowledge (such as spatial hierarchies, spatial constraints and rules for spatial qualitative reasoning) in a very elegant and natural way. Issues raised by the application of ILP to data mining are briefly discussed in the next two Subsections. It will turn out that the choice of a logic-based technique to accomplish the mining task at hand heavily affects the whole data mining process. Further details can be found in [15].

### 2.1 Data preprocessing

In SPADA, data are represented by means of Datalog atoms. For instance, the Datalog atom intersects(bari, a14) represents the fact that the highway a14 intersects the town of Bari. Thus the task of mining spatial association rules requires some pre-processing on data loaded from the spatial database. Data selection encompasses the retrieval of spatial objects eventually together with their spatial and a-spatial properties and the extraction of spatial relationships between reference objects and task-relevant objects. In particular, SPADA can extract topological relations whose semantics has been defined according to the 9-intersection model [5]. It is noteworthy that saving computational effort both in time (on-line computation of spatial relations) and in space (materialization of spatial relations) is a hot topic in spatial data mining. More sophisticated computational solutions are either the two-step spatial computation [11] or the operations on neighbourhood graphs [6]. Once selected, this data needs to be transformed in a suitable format. For instance, numerical properties of spatial objects must be discretized in order to be handled by data mining methods in the framework of computational logic. SPADA currently implements the equal-width discretization technique. An implementation of the relative unsupervised discretization algorithm RUDE [14] is ongoing.

### 2.2 From data to association rules

The task of mining spatial association rules itself can be decomposed into two sub-subtasks:
1. Find large (or frequent) spatial patterns
2. Generate highly-confident spatial association rules

The reason for such a decomposition is that frequent patterns are commonly not considered useful for presentation to the user as such. They can be efficiently post-processed into rules that exceed given threshold values. In the case of association rules the threshold values of support and confidence offer a natural way of pruning weak and rare rules. Thus, the design of algorithms for frequent pattern discovery has turned out to be a popular topic in data mining, while the generation of highly-confident rules from frequent patterns is usually performed according to the procedure given in [1].

The algorithm SPADA implements the levelwise method proposed by Mannila and Toivonen [16], which is a kind of blueprint for frequent pattern discovery algorithms. The method is based on a breadth-first search in the lattice spanned by a generality order $\leq$ between patterns. Given two patterns $P_1$ and $P_1$, we write $P_1 \leq P_2$ to denote that $P_1$ is more general than $P_2$ or equivalently that $P_2$ is more specific than $P_1$. The space is searched one level at a time, starting from the most general patterns and iterating between candidate generation and candidate evaluation phases. In SPADA, patterns are represented by means of Datalog queries, namely existentially quantified conjunctions of Datalog atoms. For instance, the Datalog query $\exists$ is_a(X, large_town)$\wedge$intersects(X, R)$\wedge$is_a(R, road) expresses an intersection pattern between a generic large town and a generic road. The pattern space is structured according to θ-subsumption [19]. For instance, consider the patterns

$$P_1 \equiv \exists\ \text{is\_a(X, large\_town)} \wedge \text{intersects(X,R)}$$

$$P_2 \equiv \exists\ \text{is\_a(X, large\_town)} \wedge \text{intersects(X, R)} \wedge \text{is\_a(R, road)}$$

We say that $P_1 \leq P_2$ because $P_2$ θ-subsumes $P_1$, namely there exists a substitution for variables ($\theta=\emptyset$)such that $P_2\theta \supseteq P_1$.

The *candidate generation* phase consists of a refinement step followed by a pruning step. The former applies the refinement operator under θ-subsumption to patterns previously found frequent while the latter mainly involves verifying that candidate patterns do not θ-subsume any infrequent pattern. In particular, a refinement step consists of adding one or more "allowed" Datalog atoms to the pattern to be refined. The list of "allowed" Datalog atoms is given by the user while defining the language of patterns. In the example above, the pattern $P_2$ is obtained by adding the Datalog atom is_a(R, road) to the pattern $P_1$. The more restrictions we put on patterns, the smaller the search space, and hence the faster a system will finish its search. The *candidate evaluation* phase is performed by comparing the support of the candidate pattern with the minimum support threshold set for the level being explored. If the pattern turns out not to be a large one, it is rejected. As for the support count, the number of distinct reference objects that satisfy the pattern is assumed as absolute frequency of the pattern in *SDB*. The support is obtained as relative frequency with respect to total number of reference objects in *SDB*.

SPADA, analogously to Geo-Associator but differently from WARMR [4], exploits is-a taxonomies for extracting multiple-level association rules. By doing so, it can explore systematically the hierarchical structure of task-relevant geographic layers. We will illustrate how SPADA works by means of an example. Let us suppose that the mining task is to discover associations relating large towns ($S$) with water bodies ($R_1$), roads ($R_2$) and province boundaries ($R_3$) in the Province of Bari, Italy. We are also given a background knowledge (BK) including the spatial hierarchies of interest. Consider that, at level $l$=2 in the spatial hierarchies, the following candidate $C$:

is_a(X, large_town), intersects(X,R), is_a(R, main_trunk_road), intersects(Y, R), diff(Y,X), is_a(Y, large_town)

has been generated after $k$=5 refinement steps and evaluated with respect to the underlying database *SDB*. Since ten of eleven large towns satisfy the pattern $C$ and all the ancestor patterns of $C$ are large at their level ($l$≤2), the pattern is a large one at level $l$=2 with support 91%. For the sake of clarity, the following pattern discovered after $k$=5 refinement steps at level $l$=1

is_a(X, large_town), intersects(X,R), is_a(R, road), intersects(Y, R), diff(Y,X), is_a(Y, large_town)

is one of the large ancestors for the pattern $C$.
Such a way of taking the taxonomies into account during the pattern discovery process implements what we referred to as the systematic exploration of the hierarchical structure of task-relevant geographic layers. Furthermore, it is noteworthy that the use of variables and the addition of the atom diff(Y,Z) derived from the BK allow the algorithm to distinguish between multiple instances of the same class of spatial objects (e.g. the class large_town). During the transformation of frequent patterns into rules, the following strong rule (91% support, 91% confidence)

is_a(X,large_town) , is_a(Y,large_town) , diff(Y,X)
→ intersects(X,R) , is_a(R,main_trunk_road) , intersects(Y,R)

has been derived from the pattern $C$. It states that "Given that 91% of large towns intersect a main trunk road which in turn is intersected by another large town distinct from the previous one, 91% of pairs of distinct large towns are crossed by the same main trunk road".

## 3. Experimentation on Stockport census data

As aforementioned, one of the main objectives of the SPIN! Project is to put to practical use the timely, cost-effective dissemination of statistical information over the internet. This will allow the project partners to evaluate the efficiency of methods, techniques and algorithms, the responsiveness of the application, as well as acceptance by customers of statistical offices.

As for the census data, the primary task was to define the application area, taking into account the advise from Eurostat to use the UK 1991 census data and to involve end users. The analysis dealt with the potential requirements of individuals and organisations for data and data processing in areas of public interest. The topic area selected for identifying potential users and their potential use of data was the public debate over Unitary

Development Plans (UDPs) in the United Kingdom. This topic area is a useful case study for several reasons:

- there are many complex problems of spatial data use
- there are similar situations in most European countries to which the methods and techniques developed for a UK case may be easily applied
- there is a very high level of recorded information about the issues and the use of data in the Public Enquiries which commonly follow the publication of plans.

The district chosen for investigation was Stockport, one of the ten Metropolitan Districts of Greater Manchester, UK. The UDP was published in 1992, subject to a public enquiry in 1995, and reviewed in 2000. It has eight main areas of policy, Environment, Countryside and Open Space, Housing and Population, Economy, Shopping, Transportation, Leisure and Community Facilities and Minerals and Waste Disposal. In order to provide a tractable subject area for analysis the Public Enquiry documentation was examined looking for areas of keenest public debate and for topics in which spatial data was central to the discussion. Two areas provided a point of focus, namely the allocation of land for Housing Development and Transport Planning.

The Stockport case study is expected to show the potential benefit of data mining methods and techniques to one or more potential users. In particular, census data are extremely important for policy analysis. Furthermore, once geo-referenced and conceptualized as spatial objects with numerical a-spatial properties, they seem to be a good test-bed for spatial association rule mining algorithms. Thus, an experimentation on Stockport census data has been designed for testing SPADA.

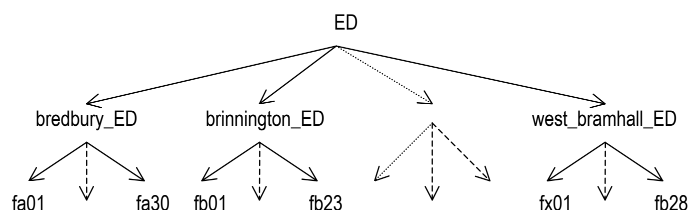


**Fig. 1.** An is-a hierarchy for the Stockport ED layer

Census data and digital ED boundaries have been loaded into an Oracle-Spatial database, i.e. a relational DBMS extended with spatial data handling facilities. The ED code has been attributed to each table as primary key in order to allow the joining of the two kinds of data and the generation of test data. We had to face several application issues mainly due to the nature of census data. Firstly, census data are all numeric. Thus attributes with a large domain have been discretized and possibly normalized. Census attributes in our experimentation are listed in Table 1. Data on employment (see attributes c24- c35) refer to persons aged 16 and over. Secondly, census data are geo-referenced with respect to the following hierarchy

$$\text{ED} \rightarrow \text{Ward} \rightarrow \text{District} \rightarrow \text{County}$$

based on the *inside* relationship among locations. The Stockport district is divided into twenty-two wards (Bredbury, Brinnington, Cale Green, Cheadle, Cheadle Hulme North, Cheadle Hulme South, Davenport, East Bramhall, Edgeley, Great Moor, Hazel Grove, Heald Green, Heaton Mersey, Heaton Moor, Manor, North Marple, North Reddish, Romiley, South Marple, South Reddish, West Bramhall), each of which consists of thirty

EDs in average. Stockport census data are actually available at the ED level. Furthermore, the current version of SPADA deals only with *is-a* hierarchies where the is-a relationship is overloaded, i.e. it may stand for *kind-of* as well as for *instance_of* depending on the context. Thus, EDs have been grouped so that EDs belonging to the same ward fall in the same class as shown in Figure 1. Further is-a hierarchies could be derived by resorting to clustering algorithms.

**Table 1.** Numerical attributes in the application to Stockport census data.

| Attr. | Description | Discretized domain |
|---|---|---|
| c24 | Total Females of Employees (full time) | Low ([1.0, 40.33[)<br>Medium ([40.33, 70.66[)<br>High ([70.66, 119.0]) |
| c25 | Total Males of Employees (full time) | Low ([0.0, 59.33[)<br>Medium ([59.33, 118.66[)<br>High ([118.66, 178.0]) |
| c26 | Total Females of Employees (part time) | Low ([0.0, 26.0[)<br>Medium ([26.0, 52.0[)<br>High ([52.0, 78.0]) |
| c27 | Total Males of Employees (part time) | Low ([0.0, 4.66[)<br>Medium ([4.66, 9.33[)<br>High ([9.33, 14.0]) |
| c28 | Total Females of Self-employed - with employees | Low ([0.0, 5.33[)<br>Medium ([5.33, 10.66[)<br>High ([10.66, 16.0]) |
| c30 | Total Males of Self-employed - with employees | Low ([0.0, 19.33[)<br>Medium ([19.33, 38.66[)<br>High ([38.66, 58.0]) |
| c32 | Total Females of On a Government scheme | Low ([0.0, 1.66[)<br>Medium ([1.66, 3.33[)<br>High ([3.33, 5.0]) |
| c33 | Total Males of On a Government scheme | Low ([0.0, 3.0[)<br>Medium ([3.0, 6.0[)<br>High ([6.0, 9.0]) |
| c34 | Total Females of Unemployed | Low ([0.0, 10.0[)<br>Medium ([10.0, 20.0[)<br>High ([20.0, 30.0]) |
| c35 | Total Males of Unemployed | Low ([0.0, 24.33[)<br>Medium ([24.33, 48.66[)<br>High ([48.66, 73.0]) |
| c36 | Total Car availability in All Households (Households with 3 or more cars counted as having 3 cars) | Low ([5.0, 151.66[)<br>Medium ([151.66, 298.33[)<br>High ([298.33, 445.0]) |

Preliminary results have been obtained by running the algorithm SPADA on such preprocessed census data. We have focused our attention on transportation planning, which is one of key issues in the UDP. An example of mining task is to discover frequent patterns relating

- EDs with a high rate of employment and car availability (*S*)

- EDs (*R*) to be characterised with respect to data about percentage of self-employment with employees.

with thresholds for the minimum support *min_sup*[1]=0.75 and *min_sup*[2]=0.05.
We are also given some background knowledge (BK) that, besides the hierarchy in Figure 1, includes the following rule:

ed_with_high_employment_and_car_availability(X) :- is_a(X,ed), c32(X,low), c33(X,low), c34(X,low), c35(X,low), c36(X,medium).

to express the constraint on *S*. From the extraction of spatial objects in *S* and *R*, the selection and transformation of numerical properties of interest (see Table 1) and the computation of the adjacency relationship among objects in *S* and objects in *R*, we obtain the Datalog database *D*(*S*) to be given in input to the system. The specification of language bias and other settings is also required. It serves as directives to the candidate generation phase.
Some interesting patterns have been discovered. For instance, at level *l*=1 in the spatial hierarchy, the following frequent pattern:

20) ed_with_high_employment_and_car_availability(A), adjacent_to(A,B), is_a(B,ed), adjacent_to(A,C), C\=B , c28(B,low) , c30(C,low)
AbsFreq 201 , Support=[0.9804878048780488]
Candidate n. 84 , Level=1 , K=6

has been generated after *k*=6 refinement steps.
By exploring systematically the is-a hierarchy on ED layer, some frequent descendants of the pattern n. 20:

267) ed_with_high_employment_and_car_availability(A), adjacent_to(A,B), is_a(B,bredbury_ED), adjacent_to(A,C) , C\=B , c28(B,low) , c30(C,low)
AbsFreq 16 , Support=[7.804878048780488e-002]
Candidate n. 3048 , Level=2 , K=6

271) ed_with_high_employment_and_car_availability(A), adjacent_to(A,B), is_a(B,cheadle_ED), adjacent_to(A,C) , C\=B , c28(B,low) , c30(C,low)
AbsFreq 17 , Support=[8.292682926829269e-002]
Candidate n. 3044 , Level=2 , K=6

275) ed_with_high_employment_and_car_availability(A), adjacent_to(A,B), is_a(B,cheadle_hulme_north_ED) , adjacent_to(A,C) , C\=B , c28(B,low) , c30(C,low)
AbsFreq 21 , Support=[0.1024390243902439]
Candidate n. 3040 , Level=2 , K=6

….

have been generated after *k*=6 refinement steps at level *l*=2.
One may wonder whether such frequent patterns convey novel knowledge and, in positive case, what kind of knowledge. The evaluation of data mining results is beyond the scope of this paper. Nevertheless a naive interpretation of results in our application might lead us to state that in the Stockport district there is no high risk of traffic congestion due to employees living in a certain ED and driving to work in adjacent EDs.

## 4. Conclusions and future work

In this paper a logic-based technique for association rule mining in spatial data has been presented. The application to census data of Manchester Stockport shows that the expressive power of computational logic enables us to tackle applications that cannot be handled by statistical spatial analysis. The experimentation is the first step towards the fulfillment of one of the main objectives of the SPIN! Project, namely to put to practical use the timely, cost-effective dissemination of statistical information over the Internet.

For the near future, we plan to optimize the system SPADA. Besides the issues of efficiency and scalability that are of great interest to data mining community, the issue of robustness (noise handling, for instance) will be faced. It is noteworthy that, in spatial data mining, robustness has another facet. Indeed, while the discovery of association rules in transactions requires little transformation of stored data, the task of mining spatial association rules relies on a more complex data pre-processing which is error-prone. For instance, the generation of the predicates close_to or adjacent_to is based on the user-defined semantics of the closeness and adjacency relations, which should necessarily be approximated. Further work on the data selection and transformation is expected to give some hints on this issue. As for the test on real-world spatial data sets, much work has also to be done. In particular, we are interested in experiments that highlight the strong interaction among census data and topographic information. Indeed, geographic layers supply topographic information which enrich the range of spatial objects and may impact on the meaningfulness of spatial relationships between EDs. For instance, once the hydrographic layer has been superimposed, it might come out that two adjacent EDs are actually separated by a river. Conversely, by taking the infrastructure layer into account, two non-adjacent EDs may turn out to be connected by a highway. These examples show that the interpretation of adjacency and closeness relations can change as spatial objects are added.

## Acknowledgements

We would like to thank Jim Petch, Keith Cole and Mohammed Islam (MIMAS, University of Manchester, England) and Chrissie Gibson (Department of Environmental and Geographical Sciences, Manchester Metropolitan University, England) for providing census data and digital ED boundaries of Manchester Stockport.

**References**


[1] Agrawal, R.; and Srikant, R. 1994. Fast Algorithms for Mining Association Rules. In *Proceedings of the Twentieth VLDB Conference*, Santiago: Cile.

[2] Ceri, S.; Gottlob, G.; and Tanca, L. 1989. What you Always Wanted to Know About Datalog (And Never Dared to Ask). *IEEE Transactions on Knowledge and Data Engineering* 1(1): 146-166.

[3] Dehaspe, L.; and De Raedt, L. 1997. Mining Association Rules in Multiple Relations. In Lavrac, N.; and Dzeroski, S. (Eds.), *Inductive Logic Programming*, LNCS 1297, Springer-Verlag, 125-132.

[4] Dehaspe, L.; and Toivonen, H. 1999. Discovery of frequent Datalog patterns. *Data Mining and Knowledge Discovery* 3(1): 7 - 36.

[5] Egenhofer, M.J.; and Herring, J.R. 1994. Categorizing Binary Topological Relations Between Regions, Lines, and Points in Geographic Databases. In Egenhofer, M.J.; Mark, D.M.; and Herring, J.R. (eds): *The 9-Intersection: Formalism and its Use for Natural-language Spatial Predicates*. Technical Report 94-1, U.S. National Center for Geographic Information and Analysis.

[6] Ester, M.; Kriegel, H.P.; and Sander, J. 1997. Spatial Data Mining: A Database Approach. In Scholl, M.; Voisard, A. (Eds.): *Advances in Spatial Databases*, LNCS 1262, Springer-Verlag, 47-66.

[7] Ester, M.; Kriegel, H.P.; and Sander, J. 1999. Knowledge Discovery in Spatial Databases. In Burgard, W.; Christaller, T.; Cremers, A.B. (Eds.): *KI-99: Advances in Artificial Intelligence*, LNCS 1701, Springer-Verlag, 61-74.

[8] Gatrell, M.R. 1991. Concepts of space and geographical data. In Maguire D.J., Goodchild M.F., Rhind D.W. (eds): *Geographical Information Systems: principles and application*. Harlow, Longman/New York, John Wiley & Sons Inc. Vol. 1: 119-134.

[9] Haining, R. 1990. *Spatial data analysis in the social and environmental sciences*. Cambridge University Press.

[10] Han, J.; Koperski, K.; and Stefanovic, N. 1997. GeoMiner: A System Prototype for Spatial Data Mining. In Peckham, J. (Ed.): *SIGMOD 1997, Proceedings of the ACM-SIGMOD International Conference on Management of Data*. SIGMOD Record 26(2), 553-556.

[11] Koperski, K.; Adhikary, J. and Han, J. 1996. Spatial Data Mining: Progress and Challenges. In *Proceedings Workshop on Research Issues on Data Mining and Knowledge Discovery*, Montreal, Canada.

[12] Koperski, K.; and Han, J. 1995. Discovery of Spatial Association Rules in Geographic Information Databases. In Egenhofer, M.J.; and Herring, J.R. (Eds.): *Advances in Spatial Databases*, LNCS 951, Springer-Verlag, 47-66.

[13] Lavrac, N.; and Dzeroski, S. 1994. *Inductive Logic Programming: Techniques and Applications*. Chichester, UK: Ellis Horwood.

[14] Ludl, M.-C.; and Widmer, G. 2000. Relative Unsupervised Discretization for Association Rule Mining. In D.A. Zighed, H.J. Komorowski, J.M. Zytkow (Eds.): *Principles of Data Mining and Knowledge Discovery*, LNCS 1910, Springer-Verlag, 148-158.

[15] Malerba, D.; Esposito, F.; and Lisi, F.A. 2001. A Logical Framework for Frequent Pattern Discovery in Spatial Data. To appear in *Proc. 14$^{th}$ International FLAIRS Conference (special track on spatiotemporal reasoning)*, Key West, Florida, USA, May 2001.

[16] Mannila, H.; and Toivonen, H. 1997. Levelwise search and borders of theories in knowledge discovery. *Data Mining and Knowledge Discovery* 1(3): 259-289.

[17] Martin, D.J. 1999. Spatial representation: the social scientist's perspective. In Longley, P.A., Goodchild, M.F., Maguire, D.J. and Rhind, D.W. (eds): *Geographical Information Systems, vol.1, Principles and Technical Issues*, 2$^{nd}$ edition, John Wiley & Sons, Inc. 71-80.

[18] Openshaw, S. 1991. Developing Appropriate Spatial Analysis Methods for GIS. In Maguire, D.J., Goodchild, M.F., Maguire, D.J. and Rhind, D.W. (eds): *Geographical Information Systems. Principles and Applications*. Longman Scientific & Technical, Burnt Mill, UK.

[19] Plotkin, G. 1970. A note on inductive generalization. *Machine Intelligence* 5: 153-163.

[20] Shaw, G., and Wheeler, D. 1994. *Statistical techniques in Geographical Analysis*. London, David Fulton.