Spatial Associative Classification:
Propositional vs Structural approach

Michelangelo Ceci   Annalisa Appice

Dipartimento di Informatica, Università degli Studi
via Orabona, 4 - 70126 Bari - Italy
{ceci, appice}@di.uniba.it

Abstract. Spatial associative classification takes advantage of employing association rules for spatial classification purposes. In this work, we investigate spatial associative classification in multi-relational data mining setting to deal with spatial objects having different properties, which are modeled by as many data tables (relations) as the number of spatial object types (layers). Spatial classification is based on two alternative approaches: a propositional approach and a structural approach. The propositional approach uses spatial association rules to construct an attribute-value representation (propositionalisation) of spatial data and performs spatial classification according to well-known propositional classification methods. Since the attribute-value representation should capture relational properties of spatial data, multi-relational association rules are used in propositionalisation step. The structural approach resorts to an extension of naïve Bayes classifiers to multi-relational data where the classification is driven by multi-relational association rules modelling regularities in spatial data. In both cases the spatial associative classification is performed at different levels of granularity and takes advantage from domain knowledge expressed in form of hierarchies and rules. Experiments on real-world geo-referenced census data analysis show the advantage of the structural approach over the propositional one.

1   Introduction

In the classical Data Mining classification setting [24] data are generated independently and with an identical unknown distribution on some domain $\mathbf{X}=X_1 \times X_2 \times \ldots \times X_m$, and are labeled according to an unknown function g whose range is a set of class labels $Y=\{C_1, C_2, \ldots, C_L\}$. A classification algorithm takes a training sample $S=\{(\mathbf{x},\ y) \in \mathbf{X} \times Y \mid y=g(\mathbf{x})\ \}$ as input and returns a function f which is hopefully close to g on the domain $\mathbf{X}$ by finding regularities in examples of a class that characterize the class in question and discriminate it from the other classes.

Classification is high demanded in Spatial Data Mining, where the training sample S is defined as a collection of (spatial) objects that are characterized by a geometrical representation (e.g. point, line, and region in 2D) as well as several non-spatial attributes. The geometrical representation and the relative positioning of spatial objects with respect to a reference system define implicitly spatial relations of different nature,

such as directional and topological. Modeling these relations is a key challenge in classification problems that arise in spatial domains [27] where data consists of multiple target spatial objects, possibly spatially-related with other non-target spatial objects. The goal of the spatial classification is to learn the concept associated with each class on the basis of the interaction of two or more spatially-referenced objects or space-dependent attributes, according to a spacing or set of arrangements [15].

The classical classification setting is restrictive for such spatial classification purposes due to *single-table* representation formalism [29], where data to be mined are represented in a single table (or relation) of a relational database. Each tuple represents an independent unit of the sample population and columns correspond to properties of units. In spatial data mining applications this assumption turns out to be a great limitation when different geographical objects have different properties, which are modeled by as many data tables as the number of spatial object types (layers).

Multi-relational data mining [11] is the branch of data mining research that overcomes these limitations and investigates methods to extract knowledge from data stored in multiple relational data tables. Multi-relational data mining methods are typically based on two alternative approaches: a propositional approach and a structural approach. In the first case, known as *propositionalisation*, a transformation of relational learning problems into classical attribute-value representations amenable for conventional data mining methods is performed. In the second case, the whole hypothesis space is directly explored by the mining method. Methods implementing structural approaches are, in principle, more powerful than methods implementing propositional approaches since information about how data were originally structured is not lost [4]. On the other hand, approaches to multi-relational data mining based on propositionalisation allow the reduction of the search space to a minimal subset including features obtained as transformation of the original multi-relational feature space. The transformation of an original multi-relational problem into a single table format allows one to directly apply conventional propositional data mining methods, thus making a wider choice of robust and well-known algorithms available [16].

In this paper, we analyse and compare the propositional and the structural approach to spatial classification in multi-relational data mining. Both approaches are investigated in the context of the associative classification framework [10, 3, 20] that combines association rules discovery and classification by taking advantage of employing association rules for classification purposes.

Several effective and efficient classifiers, e.g., CBA [20] and CMAR [18], CPAR [28], have been already built by careful selection of high quality association rules from the training data set. Such methods show that associative classification has several advantages. First, differently from most of classifiers as decision trees, association rules consider the simultaneous correspondence of values of different attributes, hence achieving better accuracy [3]. Second, it makes association rule mining applicable to classification tasks. Third, the user can decide to mine both association rules and a classification model in the same data mining process [20]. Fourth, the associative classification approach helps to solve *understandability* problems [6, 25] that may occur with some classification methods. Indeed, many rules produced by standard classification systems are difficult to understand because these systems often use only domain independent biases and heuristics, which may not fulfil user's expectation. With the associative classification approach, the problem of finding understandable rules is reduced

to a post-processing task [6]; filtering based on user-defined template may help in extracting understandable rules. Although these interesting aspects, associative classification methods do not deal with sources of complexity inherit the spatial dimension. They ignore that spatially related non-target objects may be relevant for classifying some target object. Moreover, typically they are not able to exploit any kind of *spatial taxonomic knowledge*, e.g., spatial objects described at multiple levels of granularity. For instance, UK census data can be geo-referenced with respect to the hierarchy of areal objects "ED → Ward → District → County" based on the *inside* relationship between locations. This suggests to obtain descriptions at different granularity levels to be used in *multiple-level spatial classification*.

In our proposal, we extract multi-level spatial multi-relational association rules from spatial training data by exploiting taxonomic knowledge and employ these rules to mine multi-level spatial classification models according to both a propositional approach and a structural approach. The former uses the set of extracted rules to obtain an attribute-value representation of original relational data and allows the application of classical classification algorithms. Since mining the complete set of rules may be expensive in presence of huge number of rules, rules leading to redundant attributes may be ignored for classification thus reducing the hypothesis space without affecting classification accuracy. The latter evaluates the set of multi-relational rules extracted for each level in a probabilistic framework that, given a new spatial object to be classified, estimates the posterior probability of the class. This improves traditional associative classification methods that return a categorical output and convey no information on the potential uncertainty in classification.

The paper is organized as follows. In the next Section we present the main idea behind multi-level spatial associative classification. In Section 3 we present the process of extracting multi-level spatial association rules to be used for spatial associative classification purposes. In Section 4, we present two spatial associative classification methods based on a propositional and a structural approach respectively. Finally, an experimental comparison between the propositional and the structural spatial associative classification on real-world geo-referenced census data is reported in Section 5 and some conclusions are drawn.

2   Multiple-level spatial associative classification

In a general formulation the problem of multiple-level spatial classification can be formalized as follows: *Given*: a spatial database (SDB), a set S of *target objects* tagged with a class label $y \in \{C_1, C_2, ..., C_L\}$, some sets R_k, $1 \le k \le m$, of *non-target objects*, a background knowledge BK including some *spatial hierarchies* H_k on objects in R_k, M *granularity levels* in the descriptions (1 is the highest while M is the lowest), a set of *granularity assignments* ψ_k which associate each object in H_k with a granularity level, a language bias LB that constrains the search space. *Find* multi-level spatial classifiers that predict y for target objects (with possibly UNKNOWN value for y) and related non-target objects in S.

The target objects are the main subject of the description, that is, the observation units, while the non-target objects are spatial objects that are relevant for the task in

hand and are spatially related to the former. The sets R_k typically correspond to layers (e.g., river layer in a map) of the spatial database, while hierarchies H_k define *is-a* (i.e., taxonomical) relations of spatial objects in the same layer (e.g. river *is-a* water body). Objects of each hierarchy are mapped to one or more of the *M* user-defined description granularity levels in order to deal uniformly with several hierarchies at once.

At present, only few works deal with classification models to be discovered from spatial data according to the formalization given above. Ester et al. [12] have proposed a neighbourhood graph based extension of decision trees that considers both non-spatial attributes of the classified objects and relations with neighbouring objects. However, the proposed method does not take into account hierarchical relations defined on spatial objects as well as non-spatial attributes (e.g. number of residents) of neighbouring objects. In contrast, Kopersky [15] has described an efficient method that classifies spatial objects by considering both spatial and hierarchical relations between spatial objects and takes into account non-spatial attributes for neighbouring objects. However this method assumes data stored in a single table and requires that non-spatial properties of neighboring objects be represented in aggregated form causing a consequent loss of information and a change in the units of analysis. Differently, in [22] the authors have proposed to exploit the expressive power of predicate logic to represent both represent spatial relations as well as background knowledge. For this purpose, the ILP system ATRE [23] has been integrated in INGENS, a prototypical Geographical Information System (GIS) that allows to mine classification rules for geographical objects. Training is based on a set of examples and counterexamples of geographic concepts of interest to the user (e.g., ravine or steep slopes). The first-order logic representation of the training examples is extracted from maps, although it is still controlled by the user who selects level of abstraction and/or aggregation of data.

Similarly, the discovery of spatial association rules, that is, spatial and a-spatial relationships among spatial objects, has received a lot of attention in literature. A *spatial association rule* is a rule of the form "$P \rightarrow Q$ (*s*, *c*)" such that both *P* (body) and *Q* (head) are sets of literals (items), some of which refer to (spatial) properties (e.g., area(X, [10, 19]) or relations (e.g., crosses(X,Y)), and $P \cap Q = \varnothing$. $P \cup Q$ is named *pattern*. The support *s* estimates the probability p($P \cup Q$), while the confidence *c* estimates the probability p($Q|P$). The conjunction P∧Q is called *spatial pattern*.

Koperski and Han [14] have implemented the module Geo-associator of the spatial data mining system GeoMiner that mines rules from data stored in a single table of a relational database. In contrast, in [19] the authors proposed an ILP approach to extract multi-level spatial association rules, that is, association rules involving spatial objects at different granularity levels. SPADA has been implemented as a module of the system ARES (Association Rules Extractor from Spatial data) [1].

Despite the fact that spatial association rule mining is a descriptive task, while classification of spatial objects is a predictive task, spatial association rules can be employed for classification purposes. The idea is natural since frequent spatial patterns expressing class of regularities in training spatial data can be profitably used for classification. Following this idea and, in order to meet sources of complexity of spatial data presented in section 1, we extract multi-level spatial association rules to be used for classification. In order to make rules suitable for classification we bias the structure of the rules by mining rules having only the class label in the head. Moreover, in order to filter out uninteresting rules, we impose minimal support and

confidence thresholds. In particular, for each granularity level *l*, only rules passing user defined minimal threshold value of support (*minsup*[*l*]) and confidence (*minconf*[*l*]) are used in the classification step.

In the propositional approach, the set of spatial association rules having only the class label in the head is converted in a set of boolean attributes and the result is used as input for attribute-value classification algorithms. Conversely, structural approach directly evaluates the set of multi-relational rules extracted for each level in a probabilistic framework that, given a new spatial object to be classified, estimates the posterior probability of the class.

In both cases classification models are generated by exploiting spatial association rule discovered at multiple levels of granularity. In this way, the proposed method deals with all sources of complexity inherent spatial data. The classification is performed at different levels of granularity and takes advantage from domain knowledge expressed in form of rules. Finally, it is able to handle categorical as well as numerical data through a contextual discretization method.

3   Association rules mining.

The discovery of multi-level spatial association rules aims to detect multi-level associations between target objects and non-target objects. For instance, if we intend to investigate deprivation phenomenon of some urban areas, we look for spatial association rules that relate properties of some EDs (target objects) with properties of other spatial objects, such as public transport stops (non-target objects). A spatial association rule which states that (only) "in 10% of target EDs where the percentage of households with no car is high, there is a public transport stop" can lead to the conclusion that there are some seriously deprived areas in the examined territory. An indication of the gravity of this phenomenon is given by the support of the rule: if the percentage is high (e.g., 80%), then there are many target EDs with a high percentage of no-car-owning households which are not served by public transportation.

Spatial association rules can be represented in multi-relational setting by resorting to subsets of first-order logic. For instance, the association rule in the above example can be easily expressed as "*is_a(X,ed)* ∧ *no-car-owning-households%(X,high)* → *contains(X,Y)* ∧ *is_a(Y, public_transp_stop) (80%, 10%)*". The corresponding multi-relational pattern "*is_a(X, large_town)* ∧ *intersects(X,Y)* ∧ *is_a(Y, road)* ∧ *intersects(X, Z)* ∧ *is_a(Z, road)* ∧ *Z≠Y*" expresses the spatial containment relationship between a target ED and some non-target spatial object classified as public transport stop. Items are first-order logic *atoms*, i.e., *n*-ary *predicates* applied to *n terms*. In the above example terms are *variables* (*X* and *Y*) and *constants* (*road*).

By taking into account some taxonomic knowledge (*hierarchies*) on non-target objects we may obtain descriptions at different granularity levels (*multiple-level association rules*). For instance, a finer-grained association rules can be "*is_a(X,ed)* ∧ *no-car-owning-households%(X,high)* → *contains(X,Y)* ∧ *is_a(Y, bus_stop) (70%, 11%)*" which provides more insight into the non-target object *Y*, according to the hierarchy describing a *bus_stop* as specialization of a *public_transp_stop*.

Multi-level association rules discovery has been investigated in the seminal work of Han and Fu [17]. Anyway most of works deal with data represented according to single table formalism. The ILP system WARMR [8] overcomes this limitation by discovering frequent DATALOG patterns (multi-relational patterns). Although WARMR has been presented as a system able to mine taxonomical knowledge, it lacks of mechanisms for taxonomic reasoning to mine multi-level association rules. Moreover, the transformation of frequent patterns into rules requires the specification of a list of patterns that can occur in the head or body of a rule. Therefore, the bias for the output rules is a must and not an additional opportunity given to filter some rules. Finally, WARMR is not integrated in a system for spatial data analysis that supports users to extract (spatial) properties of either target objects or non-target object.

Therefore, to the best of our knowledge, SPADA can be considered the only multi-relational data mining method especially conceived for extracting multi-level spatial association rules involving relations among spatial objects stored in a spatial database (SDB). SPADA operates in three steps for each granularity level: i) pattern generation; ii) pattern evaluation; iii) rule generation and evaluation. SPADA takes advantage of statistics computed at granularity level l when computing the supports of patterns at granularity level $l+1$. Further details are provided in [19].

SPADA is integrated in ARES (http://www.di.uniba.it/~malerba/software/ARES/index.htm) a spatial data mining system that is loosely coupled with a Oracle Spatial 9i SDB. Data stored in the SDB are pre-processed and then represented in a deductive database (DDB). A middle layer makes possible a loose coupling between SPADA and the SDB by generating features of spatial objects. This middle layer includes both the module RUDE (Relative Unsupervised DiscrEtization) to discretize a numerical attribute of a relational database in the context defined by other attributes [21] and the module FEATEX (Feature Extractor) that is implemented as an Oracle package of procedures and functions, each of which computes a different feature. According to their nature, features extracted by FEATEX are distinguished as geometrical (e.g. area and length), directional (e.g. direction) and topological features (e.g. intersects) [1]. Extracted features are represented by extensional predicates. For instance, spatial intersection between two objects X and Y is expressed with $intersects(X,Y)$. In this way, the expressive power of first-order logic in databases is exploited to specify both a background knowledge BK including spatial hierarchies and domain knowledge, and the language bias LB. Spatial hierarchies allow to represent spatial objects at different levels of granularity, while the domain knowledge stored as a set of rules in the intensional part of the DDB supports qualitative reasoning. The LB allows data miners to specify bias for interesting solutions, and then to exploit this bias to improve both the efficiency of the mining process and the quality of the discovered rules. In SPADA, the LB is a set of constraint specifications for either patterns or association rules. Pattern constraints allow to specify a literal or a set of literals that should occur one or more times in discovered patterns. During the rule generation phase, patterns that do not satisfy a pattern constraint are filtered out. This means that they are generated and evaluated anyway. This late exploitation of pattern constraints is due to the fact that if a pattern P does not satisfy a constraint, it is still possible that P descendants (i.e., more specific patterns) satisfy it. Therefore, pattern constraints do not prune the pattern space, but improve the efficiency of the mining process, since they prevent the generation and evaluation of useless rules. Similarly, rule constraints

are used to specify literals that should occur in the head or body of discovered rules. Finally a pattern (rule) constraint is used to specify both minimum and maximum number of occurrences for a literal in a pattern (head or body of a rule). In this way users may define the head structure of a rule requiring the presence of exactly a specific literal and nothing more. Consequently, we are able to constraint the head structure of a rule requiring the presence of only the class label.

4   Classification step

Once spatial association rules having only the class label in the head are extracted for each level, they are used in classification step. In the following subsections we show how to exploit such rules to mine both propositional and structural spatial classifiers.

Propositional Approach

In our proposal, the propositional classifier is built by converting rules in a set of boolean features and using the result as input for attribute-value classification algorithms. More precisely, for each granularity level l, boolean features construction is performed by transforming spatial data stored in multiple relational tables of a relational Database in a single relational table B whose columns correspond to each body of association rules discovered at level l. Before describing how B is obtained, we introduce: *i)* R': a set of relational association rules whose head contains a literal representing the class label; *ii)* I_i a target spatial object; *iii)* B: the output table with $|R'|+1$ attributes. Tuples in B correspond to target spatial objects to be classified.

The instance I_i can be logically represented as a set of ground facts describing both the target object and all non-target objects that are (spatially) related to the target object. The body of a rule $R_j \in R$' *covers* I_i if there exists a substitution θ, such that $body(R_j)\theta \subseteq I_i$. In the case that the body of the rule R_j covers I_i, the j-th value of the tuple in B associated to the i-th instance of T is *true*, otherwise it is *false* (see Example 1). The $(|R'|+1)$-th column in B represents the target attribute y.

Example 1: The multi-relational description of the instance "*is_a(e, ed), y(e, high_deprivation), inhabitants(e, 2000), contains(i, s_1), is_a(s_1, school), students(s_1, 105), contains(e, s_2), is_a(s_2, shop), employed(s_2, 3), contains(e, s_3), is_a(s_3, shop), employed(s_3, 1), distance(s_1, s_2, 20), …*" describing the target ED e and the spatially related non-target objects s_1, s_2 and s_3 can be converted into *(true, false)* attribute-value representation according to the rules :

R1: *is_a*(X, *ed*) $\land$ *inhabitants*(X,[1000,3000]) $\land$ *inhabitants*(X,[1000,3000]) $\land$ *contains*(X, Y), *is_a*(Y, shop) $\land$ *employed*(Y,[1,2])$\rightarrow$ *y*(X, high_deprivation) .

R2: *is_a*(X, *ed*) $\land$ *inhabitants*(X,[1000,3000]) $\land$ *inhabitants*(X,[1000,3000]) $\land$ *contains*(X, Y), *is_a*(Y, school) $\land$ *students*(Y,[50,200])$\rightarrow$ *y*(X, low_deprivation) .

It is noteworthy that the number of attributes of B depends on the number of discovered association rules in R'. Since, the number of discovered association rules is usually high and many rules are strongly correlated, this may lead to generate boolean features which are highly correlated or even logically redundant.

In this work, we assume that boolean features (attributes) are redundant if their removal does not change the set of example-pairs having the same value for each feature. A feature p is identified as *redundant* with respect to another feature q for

distinguishing *positive* from *negative* examples of a class *c* if *p* is true for at least the same positive examples as *q*, and false for at least the same negative examples as *q*. Coherently with this definition, the association rule based propositionalisation is combined with redundant feature elimination to reduce the hypothesis space by excluding boolean features which are redundant for learning. Feature reduction is performed by integrating the feature reduction method REFER [2] that operates in ILP setting for multi-class problems. REFER preserves the existence of a complete and consistent theory for each class when eliminating redundant features [2]. This method appears well-suited to reduce the set of propositionalized features at each granularity level, improving efficiency of classification without comprising the existence of some classification model discriminating between examples tagged with different class label.

The learning process takes as input the reduced relational table and returns a classification model built by propositional learning algorithms as JRIP [7], C4.5 [28], 1-NN [24] and the Naive Bayesian classifier (NBC) [9]. JRIP is a rule learner similar to the commercial rule learner Ripper, C4.5 is a well known decision tree learner, 1-NN is an instance-based learner that that assigns the class of the closest training example to a new one, Naive Bayesian classifier is another well-known classifier based on the bayes theorem for conditional probability. In Figure 1, the flow diagram of the multi-relational propositionalization-based associative classifier is shown.

Structural Approach

Concerning the structural approach, the set of rules is used in the construction of a naïve Bayesian classifier [9], which aims to classify any target object *o* by maximizing the *posterior probability* $P(C_i|o)$ that *o* is of class C_i, that is $class(o) = arg\ max_i\ P(C_i|o)$. By applying the Bayes theorem, $P(C_i|o)$ can be reformulated as:

$$P(C_i|o) = \frac{P(C_i)P(o|C_i)}{P(o)} . \qquad (1)$$

The term $P(o|C_i)$ is estimated by means of the *naïve Bayes assumption*: $P(o|C_i) = P(o_1, o_2,\dots, o_m|C_i) = P(o_i|C_i) \times P(o_2|C_i) \times \dots \times P(o_m|C_i)$ where o_1, o_2, ..., o_m represent the set of the properties, different from the class, used to describe the object. This assumption is clearly false if the predictor variables are statistically dependent. However, even in this case, the naïve Bayesian classifier can give good results [9]. In (1) the value $P(C_i)$ is the prior probability of the class C_i. Since $P(o)$ is independent of the class C_i, it does not affect $f(o)$, that is $class(o) = arg\ max_i\ P(C_i)P(o|C_i)$.

This formulation of the problem holds under the single-table assumption. In order to mine relations of a target object, the problem formulation must be extended by introducing multi-relational (spatial) rules to guide the computation of $P(o|C_i)$. Given the target spatial object *o*, we consider the subset of the extracted rules that can be used to classify *o*. More formally, we consider the subset *R* of rules whose body is satisfied by the object to be classified both in terms of the values of properties of involved spatial objects and in terms of the spatial relations between objects. For
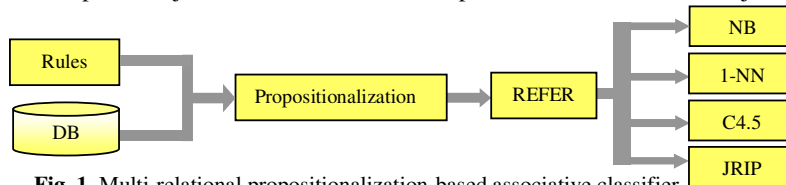

Fig. 1. Multi-relational propositionalization-based associative classifier.

example, if the set of target spatial objects is the set of wards in a district, a ward *w* satisfies the rule "*wards_relatedTo_waters(A, B),waters_typewater(B, river), cars_per_person(A, high)* → *mortality_rate(A, low)* " when *w* is spatially related (intersects) to a river and is characterized by a high average number of cars per person. We use *R* to estimate *P(o|C$_i$)*. In particular, we estimate *P(o|C$_i$)* by means of the probabilities associated to both spatial relations (e.g. *wards_relatedTo_waters(A,B)*) and properties (e.g. *waters__typewater(B,RIVER)*, *cars_per_person(A,high)*) associated to each rule in *R*:

$$P(o \mid C_i) = \prod_{k \in |R|} (P(relations_k \mid C_i) \prod_j P(property_{k,j} \mid relations_k, C_i)) \qquad (2)$$

where the term *relations$_k$* represents the event that the set of spatial relations expressed in the *k*-th rule is satisfied, while the term *property$_{k,j}$* represents the event that the *j*-th property of the *k*-th rule is satisfied. Further details are reported in [5].

5   Experiments

In this section we present the comparison of the propositional and the structural spatial associative classifiers in geo-referenced census data interpretation. For this study we consider both census and digital map data provided in the European project SPIN! (http://www.ais.fraunhofer.de/KD/SPIN/project.html). These data concern Greater Manchester, one of the five counties of North West England (NWE). Greater Manchester is divided into ten metropolitan districts, each of which is decomposed into censual sections (wards), for a total of two hundreds and fourteen wards. Census data are available at ward level and provide socio-economic statistics (e.g. mortality rate, i.e., the percentage of deaths with respect to the number of inhabitants) as well as measures describing the deprivation of each ward according to information provided by Census combined into single index scores. We consider Jarman Underprivileged Area Score that is designed to measure the need for primary care, the indices developed by Townsend and Carstairs that is used in health-related analyses, and the Department of the Environment's Index (DoE) that is used in targeting urban regeneration funds. The higher the index value the more deprived a ward is. Both deprivation indices values as well as mortality rate are all numeric, but association rules discovery deals with discrete values automatically discretized by means of RUDE. Jarman index, Townsend index, DoE index and Mortality rate have been discretized in (*low, high*), while Carstairs index has been discretized in (*low, medium, high*).

By considering Greater Manchester wards as target objects, we focus our attention on mining a classification model to predict discrete value of DoE index by exploiting not only socio-economic census factors but also geographical factors represented in some linked topographic maps. Spatial analysis is enabled by the availability of vectorized boundaries of the 1998 census wards as well as by other Ordnance Survey digital maps of NWE, where several interesting layers such as road net (1687 lines), rail net (805 lines), water net (716 lines, urban area (115 lines) and green area (9 lines) are found. These additional layers are non-target objects of our analysis. Both ward-referenced census data and map data are stored in an Object-Relational spatial database, i.e., Oracle Spatial 9i database, as a set of spatial tables, one for each layer.

Each spatial table includes a geometry attribute to store the geometrical representation and the positioning of a spatial object with respect to a reference system.

Mining this geometrical data poses the problem of extracting implicit topological relationships between target objects and non-target objects. At this purpose, we exploited FEATEX *relate* function to verify the existence of some topological relationship (i.e., intersects, inside, connects) between *ward_135*, that is a specific Greater Manchester ward, and *urbareaL_151*, that is a large urban area. In these experiments, the number of computed "non disjoint" relationships is 5313 (13 wards-green areas, 2798 wards-roads, 1067 wards-waters, 381 wards-urban areas and 1054 wards-rails). To support a spatial qualitative reasoning, a domain specific knowledge has been expressed in the *BK* as a set of rules. Some of these rules are:

> *crossed_by_urbanarea(X,Y) :- connects(X,Y), is_a(Y, urban_area).* …
> *crossed_by_urbanarea(X,Y) :- inside(X,Y), is_a(Y, urban_area).*

To mine association rules at multiple level of granularity, we enrich *BK* with five hierarchies on the non-target objects (see Figure 2). They have depth three and are straightforwardly mapped into three granularity levels.

Performances of the two approaches are evaluated by means of a 10-fold cross validation. The first step consists in the extraction of association rules having class label in the head. For this purpose, we specify a language bias (*LB*) to constrain the search space and to filter out uninteresting spatial association rules. In particular, we rule out all spatial relationships (e.g. connects, inside, and so on) directly extracted by FEATEX and asked for rules containing topological predicates defined by means of *BK* rules (i.e., crossed_by_urbanarea, crossed_by_road_net, crossed_by_rail_net, crossed_by_green_area and crossed_by_water_net). Association rules are then discovered at different levels of granularity according to hierarchies defined on non-target objects. These rules contain useful information about spatial patterns frequently occurring in data. For instance, by analyzing spatial association rules extracted with parameters *minsup* = 0.1, *minconf* = 0.6 we discover the rule "*is_a(A, ward), ward_urbanarea(A, B), is_a(B, urban_area), jarman (A, low) → doe (A, low) (52.6%, 100%)*" which states that a low DoE index value is observed in a ward *A* that includes an urban area *B* and has a low value of Jarman index. The support (52.6%) and the high confidence (100%) confirm a meaningful association between a geographical factor, such as living in urban areas where primary care are well satisfied, and low level of derivation when considering targeting urban regeneration funds.

At a granularity level 2, SPADA specializes the non-target object *B* by generating the following rule that preserves both support and confidence "*is_a*(*A, ward*), *ward_urbanarea*(*A, B*), is_a*(*B,* large_urban_area*), jarman *(A, low) → doe (A, low) (52.6%, 100%)*". This rule clarifies that the urban area *B* is large. By varying granularity level as well as value of *minsup*, *minconf* and number of refinement steps *K* (pattern length) in association rule discovery, we obtain several experimental settings.

The second step consists in using association rules discovered at each granularity level to mine a classification model. In particular, concerning the propositional
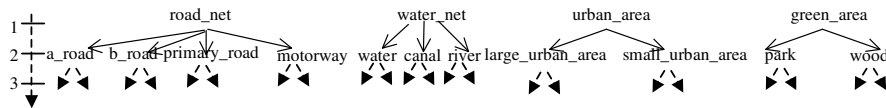

Fig. 2. Spatial hierarchies defined for road net, water net, urban area and green area.

approach, the set of discovered rules is transformed in a set of boolean features that is the conjuctions of literals derived from the body of each rule, and redundant features are removed to reduce feature space without affecting the existence of a complete and consistent theory for each class label. In the previous example, the feature construction step generates the boolean feature "*is_a(A, ward)* ∧ *ward_urbanarea(A, B)* ∧ *is_a(B, urban_area)* ∧ *jarman(A, low)*" that is true for each Greater Manchester ward with a low value of Jarman index such that there is at least an urban area that is topologically related (e.g. connects or inside) to the ward in question. Conversely, in structural approach, the entire set of association rules discovered at each granularity level contributes to estimate the posterior probability as explained in Section 3.

Table 1. Feature reduction factor (propositional approach).

min sup	min conf	Granularity level	K	Avg. number of original features	Avg. perc. of feature reduction
0.2	0.8		5	27.5	37%
0.2	0.8	1	6	221.1	44%
0.2	0.8		7	982.3	73%
0.2	0.8		5	492.8	29%
0.2	0.8	2	6	278.3	11%
0.2	0.8		7	1860.1	66%
0.1	0.6		5	233.2	48%
0.1	0.6	1	6	374	57%
0.1	0.6		7	904.2	74%
0.1	0.6		5	442.2	28%
0.1	0.6	2	6	635.8	27%
0.1	0.6		7	3051.4	68%

The results on the average feature reduction percentage are reported in Table 1. They confirm that high number of association rules typically lead to redundant features. Percentage of feature reduction increases when the refinement step increases. This is due to the high number of similar rules (thus producing redundant features).

Table 2. DoE Index average accuracy

min sup	min conf	Granularity level	K	Avg. Accuracy				
				Propositional approach				Structural approach
				NBC	1-NN	C4.5	JRIP	
0.2	0.8		5	81.64	80.73	*83.01*	*83.01*	**87.5**
0.2	0.8	1	6	80.28	80.73	*83.01*	*83.01*	**87.5**
0.2	0.8		7	75.56	80.28	***83.01***	***83.01***	82.1
0.2	0.8		5	82.55	82.55	81.64	81.93	**90.28**
0.2	0.8	2	6	82.10	82.55	82.10	*83.01*	**88.3**
0.2	0.8		7	80.28	82.55	81.19	81.02	**87.4**
0.1	0.6		5	83.01	80.73	*83.01*	*83.01*	**91.2**
0.1	0.6	1	6	80.28	80.73	*83.01*	*83.01*	**91.2**
0.1	0.6		7	80.73	80.73	*83.01*	*83.01*	**91.2**
0.1	0.6		5	83.01	81.30	82.10	81.02	**91.2**
0.1	0.6	2	6	82.55	79.65	80.28	81.64	**91.2**
0.1	0.6		7	80.73	79.31	*81.19*	81.02	**91.2**

Table 3. DoE Index average running times (secs). Results are obtained on a Intel Centrino (1.3 GHz) PC running WinXP

min sup	min conf	Granularity level	K	Proposition alization step	Feature reduction step	Classification				
						Propositional approach				Structural approach
						NBC	1-NN	C4.5	JRIP	
0.2	0.8		5	12.7	0.017	0.01	0.011	0.03	0.046	20.1
0.2	0.8	1	6	10.1	0.016	0.011	0.01	0.029	0.041	28.3
0.2	0.8		7	40.7	0.03	0.01	0.027	0.035	0.047	78.0
0.2	0.8		5	25.1	0.025	0.01	0.018	0.049	0.066	56.3
0.2	0.8	2	6	24.0	0.018	0.01	0.014	0.043	0.064	56.8
0.2	0.8		7	102.5	0.041	0.013	0.03	0.064	0.075	343.8
0.1	0.6		5	11.3	0.016	0.01	0.015	0.029	0.041	21.5
0.1	0.6	1	6	16.2	0.022	0.017	0.011	0.033	0.046	35.3
0.1	0.6		7	40.6	0.0281	0.011	0.013	0.032	0.045	105.6
0.1	0.6		5	21.6	0.023	0.011	0.024	0.028	0.071	81.3
0.1	0.6	2	6	27.9	0.0291	0.013	0.02	0.048	0.128	120.4
0.1	0.6		7	160.7	0.0641	0.024	0.046	0.087	0.12	757.2

The average accuracy and running time of classification performed according to both the propositional and the structural approach are reported in Table 2 and Table 3 respectively. Classification is evaluated by varying *minsup*, *minconf* and *K* for each setting. Several considerations can be drawn.

Firstly, both the propositional and the structural Bayesian associative classifier significantly improve accuracy of the trivial classifier that returns the most probable class (acc. 0.625). Secondly, the average accuracies of classification models discovered at higher granularity levels (i.e. level = 2) are sometimes better than the corresponding accuracies at lowest levels. This means that the classification model may take advantage of the use of the hierarchies defined on spatial objects. In this case, results at different abstraction levels provide insights on what are the non-target objects that affect the classification. For instance, when *K* = 7, *minsup* = 0.2 and *minconf* = 0.8, both propositional naïve bayes associative classification and structural classifier are strongly improved when considering the size of the urban area, the type of road, and so on. Thirdly, results show that by decreasing the number of extracted rules (higher support and confidence) we have lower accuracy. This means that there are association rules that strongly influence classification results and often these rules are not characterized by high values of support and confidence. Fourthly, we observe that, in this specific task, the higher the number of refinement steps (rules involving more literals) not necessarily means the better the model. This is mainly due to the fact that rules become very specific. Fifthly, it is noteworthy that by considering different levels of granularity it is possible to find the best trade-off between Bias and Variance of the classifier [13] where Bias denotes the lack of fitting the data, while Variance is the probability of assigning a class that is different from the Bayes optimal one. In particular, at higher levels of granularity we have higher Bias and lower Variance. On the contrary, at lower levels of granularity we have lower Bias and higher Variance. Finally, results clearly show that the structural associative classifier outperforms the propositional one. This confirms that methods implementing structural approaches are more powerful than methods implementing

propositional approaches since information about how data were originally structured is not lost. On the other hand, our propositional approach outperforms the structural one when considering efficiency. This depends on the reduction of the search space in the mining process. As future work we intend to report a complexity analysis in order to formally prove the advantage of the propositional approach in term of efficiency.

6   Conclusions

In this paper, we have compared the propositional and the structural approach to spatial classification in multi-relational data mining. Both approaches are investigated in the context of the associative classification framework that combines spatial association rules discovery and classification by taking advantage of employing association rules for classification purposes. In particular, association rules that contain only the class label in the head are discovered at multiple levels of granularity. In the propositional approach, they are transformed in a set of boolean features to be used in the classical attribute-classification. Differently, in structural approach, this set of rules contributes to estimate the posterior probability.

Results show that the use of different levels of granularity permits to find the best trade-off between Bias and Variance in spatial classification. Moreover, they confirm that spatial data mining applications benefit of preserving information about how data were originally structured in spatial domain. This justifies the better accuracy of the structural approach with respect to the propositional one. On the contrary, the propositional approach outperforms the structural one in terms of efficiency due to the reduction of the search space.

Acknowledgments

References

1. Appice, A., Ceci, M., Lanza, A., Lisi, F.A. Malerba, D.: Discovery of Spatial Association Rules in Georeferenced Census Data: A Relational Mining Approach, Intelligent Data Analysis. Special issue of Mining Official Data. 7(6) (2003).
2. Appice, A., Ceci, M. Rawles, S., Flach, P.: Redundant Feature Elimination for Multi-Class Problems. *Proc. of Int'l Conf. on Machine Learning*, Morgan Kaufmann , (2004), 33-40.
3. Baralis, E., Garza, P.: Majority Classification by Means of Association Rules. In N. Lavrac, D. Gamberger, L. Todorovski & H. Blockeel (Eds.). *Proc. of Europ. Conf. on Principles and Practice of Knowledge Discovery in Databases*, LNAI 2838, (2003), 35-46.
4. Ceci, M., Appice, A. Malerba, D.: Mr-SBC: a Multi-Relational Naive Bayes Classifier. In N. Lavrac, D. Gamberger, L. Todorovski & H. Blockeel (Eds.). *Proc. of Europ. Conf. on Principles and Practice of Knowledge Discovery in Databases*, LNAI 2838, (2003), 95-106.
5. Ceci, M., Appice, A., Malerba, D.: Spatial Associative Classification at Different Levels of Granularity: A Probabilistic Approach. In J.F. Boulicaut, Esposito F., Giannoti F. & D.

Pedreschi (Eds.). *Proceedings of European Conference on Principles and Practice of Knowledge Discovery in Databases,* LNAI 3202, Springer-V., (2004), 99-111.

6. Clark, P., Matwin, S.: Using qualitative models to guide induction learning. *Proc. of Int'l Conf. of Machine Learning*, Morgan Kaufmann, (1993), 49-56.
7. Cohen, W. W.: Fast Effective Rule Induction. *Proc. of Int'l Conf. on Machine Learning,* Morgan Kaufmann (1995), 115-123.
8. Dehaspe, H. Toivonen: Discovery of frequent Datalog patterns. *Data Mining and Knowledge Discovery*, 3(1), (1999), 7-36.
9. Domingos, P. Pazzani, M.: On the Optimality of the Simple Bayesian Classifier under Zero-One Loss. *Machine Learning*, 29 (2-3), (1997), 103-130.
10. Dong, G., Zhang, X., Wong, L., Li, J.: Classification by aggregating emerging patterns. *Proc. of DS'99*, LNAI 1721, (1999).
11. Dzeroski, S., Lavrac, N. (eds.): *Relational Data Mining*. Springer-Verlag, Berlin, (2001).
12. Ester, M. Kriegel, H.P., Sander J.: Spatial Data Mining: A Database Approach. *Proc.s of Int'l Symposium on Large Databases*, (1997), 47-66.
13. German, S., Bienenenstock, E., Doursat, R.: Neural networks and the bias/variance dilemma. *Neural Computation*, 4, (1992), 1-58.
14. Koperski, K., Han, J.: Discovery of Spatial Association Rules in Geographic Information Databases. Advances in Spatial Databases. LNAI 951, (1995) 47-66.
15. Koperski, K.: Progressive Refinement Approach to Spatial Data Mining, Ph.D. thesis, Computing Science, Simon Fraser University, (1999).
16. Krogel, M., Rawles, S., Zelezny, F., Flach, P., Lavrac, N., Wrobel S.: Comparative evaluation of approaches to propositionalization. *Proc. of Int'l Conf. on Inductive Logic Programming,* LNAI 2835, . (2003), 197-214.
17. Han, J., Fu, Y.: Discovery of multiple-level association rules from large databases., *Proc. of Int'l Con. on Very Large Data Bases,* (1995), 420-431
18. Li, W., Han, J., Pei, J.: CMAR: Accurate and efficient classification based on multiple class-association rules. *Proc. of Int'l Conf. on Data Mining,* (2001).
19. Lisi, F.A., Malerba, D.: Inducing Multi-Level Association Rules from Multiple Relations. *Machine Learning*, 55, (2004), 175-210.
20. Liu, B., Hsu, W., Ma, Y.: Integrative classification and association rule mining. *Proc. of Knowledge Discovery in Databases*, (1998).
21. Ludl, M.C., Widmer, G.: Relative Unsupervised Discretization for Association Rule Mining. In D. Zighed, H.J. Komorowski & J.M. Zytkow (Eds.). *Proc. of Europ. Conf. on Principles and Practice of Knowledge Discovery in Databases*, LNAI 1910, (2000), 148-158.
22. Malerba, D., Esposito, F., Lanza, A., Lisi, F.A., Appice, A.: Empowering a GIS with Inductive Learning Capabilities: The Case of INGENS. Journal of Computers, Environment and Urban Systems, Elsevier Science, 27 (2003), 265-281.
23. Malerba, D.: Learning Recursive Theories in the Normal ILP Setting, Fundamenta Informaticae, 57, 1, (2003), 39-77.
24. Mitchell, T. *Machine Learning*. McGraw Hill, (1997).
25. Pazzani, M., Mani, S., Shankle, W.R.: Beyond concise and colorful: learning intelligible rules. *Proc. of Knowledge Discovery in Databases,* (1997).
26. Quinlan, J.: *C4.5: Programs for machine learning*. Morgan Kaufmann, (1993).
27. Shekhar, S., Schrater, P.R., Vatsavai, R. R., Wu, W., Chawla, S.: Spatial Contextual Classification and Prediction Models for Mining Geospatial Data. *IEEE Transaction on Multimedia*, 4(2), (2002), 174-188.
28. Yin, X., Han, J.: CPAR: Classification based on predictive association rules. *Proc. of SIAM International Conference on Data Mining*, SIAM, (2003).
29. Wrobel, S.: Inductive logic programming for knowledge discovery in databases. In Džeroski, S., N. Lavrač (Eds.). *Relational Data Mining*. Springer-Verlag, (2001), 74-101.