Towards Elimination of Well Known Geographic Patterns in Spatial Association Rule Mining

Vania Bogorny, Sandro da Silva Camargo, Paulo Martins Engel, Luis Otavio Alvares

Abstract – Many spatial association rule mining algorithms have been developed to extract interesting patterns from large geographic databases. However, a large amount of knowledge explicitly represented in geographic database schemas has not been used to reduce the number of association rules. A significant number of well known dependences, explicitly represented by the database designer, are unnecessarily extracted by association rule mining algorithms. The result is the generation of hundreds or thousands of well known spatial association rules. This paper presents an approach for mining spatial association rules where both database and schema are considered. We propose the APRIORI-KC (Apriori Knowledge Constraints) algorithm to eliminate all associations explicitly represented in geographic database schemas. Experiments show a very significant reduction of the number of rules and the elimination of well known rules.

Index Terms - Geographic databases, geographic domain knowledge, spatial association rules, spatial data mining.

I. INTRODUCTION

The association rule mining technique emerged with the objective to find novel, useful, and interesting associations, *hidden* among itemsets [1] and spatial predicate sets [2]. An enormous amount of algorithms with different thresholds for reducing the number of rules has been proposed. However, only the data by themselves have been considered, while the database schema, which is a rich knowledge resource, has not been used as prior knowledge to eliminate well known patterns.

In traditional association rule mining the schema might not be useful, since items and transactions can be stored in a single relation. In geographic databases, however, the number of object types to be considered for mining is large. Every different object type is normally stored in a different relation, since most geographic databases follow the relational approach [3]. Fig. 1 shows an example of how geographic data are stored in relational databases. There is a different relation/table for every different object type [3] (street, water resource, gas station, and island).

(a) Street

Gid	Name	Length	Shape
1	BR-101	632056.03	Multiline [(x_1,y_1),(x_2,y_2),..]
2	RS-226	255365.88	Multiline [(x_1,y_1),(x_2,y_2),..]

(b) WaterResource

Gid	Name	Length	Shape
1	Jacui	3214328.71	Multiline [(x_1,y_1),(x_2,y_2),..]
2	Guaiba	283434.23	Multiline [(x_1,y_1),(x_2,y_2),..]
3	Uruguai	4523333.12	Multiline [(x_1,y_1),(x_2,y_2),..]

(c) GasStation

Gid	Name	Vol_Diesel	Vol_Gas	Shape
1	Posto do Beto	20000	85000	Point[(x_1,y_1)]
2	Posto da Silva	30000	95000	Point[(x_1,y_1)]
3	Posto Ipiranga	25000	120000	Point[(x_1,y_1)]

(d) Island

Gid	Name	Population	Sanitary_Condition	Shape
1	Flores	5000	Yes	Point[(x_1,y_1)]
2	Pintada	20000	Partial	Point[(x_1,y_1)]
3	Da Luz	15000	No	Point[(x_1,y_1)]

Fig. 1. Examples of geographic data storage in relational databases

From the database design point of view, the objective of data modeling is to bring together all relevant object types of the application, their associations/relationships, and their constraints [3]-[4]. Many geographic object types have mandatory associations, represented in the schema by *one-one* and *one-many* cardinality constraints, which the database designer has the responsibility to warrant when the schema is conceived [4]. The representation is usually in the third normal form [4], intending to reduce anomalies and warrant integrity.

In contrast to database schema modeling, where associations between data are *explicitly* represented, association rule mining algorithms should find *implicit* and novel associations. While the former represents the data into the third normal form, the latter usually denormalizes the data in one single table or one single file. This transformation brings the associations explicitly represented in the database schema to the dataset to be mined, and by consequence, many well known associations specified in the schema, are extracted by association rule mining algorithms.

In geographic databases, the number of associations specified in the schema reflects a large number of well known geographic dependences. Fig. 2 shows two layers of

This work was supported in part by the Brazilian Government (CAPES and CNPQ).

V. Bogorny is a PhD candidate in Computer Science at Instituto de Informática – UFRGS, Porto Alegre, Brazil (telephone: 55-51-3316-7013, e-mail: vbogorny @inf.ufrgs.br).

S. S. Camargo is a PhD student in Computer Science at Instituto de Informática – UFRGS, Porto Alegre, Brazil (e-mail: scamargo@inf.ufrgs.br).

P. M. Engel is a professor in Computer Science at Instituto de Informática – UFRGS, Porto Alegre, Brazil (telephone: 55-51-3316-6829, e-mail: engel @inf.ufrgs.br).

L.O. Alvares is a professor in Computer Science at Instituto de Informática – UFRGS, Porto Alegre, Brazil (telephone: 55-51-3316-6843, e-mail: alvares @inf.ufrgs.br).

information of the same geographic region. On the left there is a well known pattern, i.e., a geographic dependence where gas stations do always intersect streets. If considered in association rule mining, such dependence will produce high confidence rules (e.g. *is_a(GasStation)*$\rightarrow$ *intersect(Street)* (100%)). On the right, however, there is no explicit pattern among gas stations and water resources which may produce well known rules. Relationships such as the example shown in Fig. 2 (right) may be interesting for association rule mining.

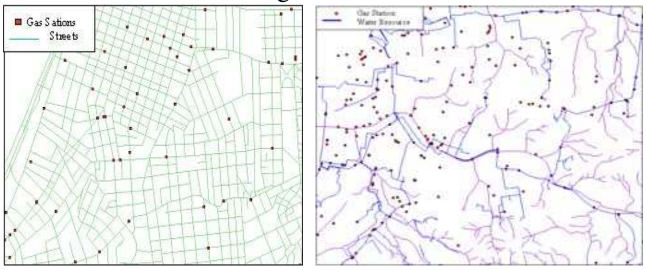


Fig. 2. Explicit well known geographic domain associations (left) and implicit spatial associations (right)

Users of some domains may not be interested in strong geographic domain rules such as *is_a(GasStation)*$\rightarrow$ *intersect(Street)* (100%), but in non-obvious rules such as *is_a(GasStation) and intersect(WaterResource)* $\rightarrow$ *pollution=high* (70%).

In geographic databases, most mined rules are strongly related to geographic dependences which represent strong regularities, but do not contribute to the discovery of novel and useful knowledge. The result is the mixed presentation of thousands of interesting and uninteresting associations that can discourage users from interpreting them all in order to find interesting patterns.

We claim that well known associations, explicitly represented in geographic database schemas, should be eliminated in association rule mining to avoid their extraction and presentation to the user. Although well know associations can be reduced in geographic data preprocessing steps [5], most dependences can only be eliminated into the data mining algorithm. Aiming to reduce the amount of well known patterns in spatial association rule mining, this paper proposes the APRIORI-KC algorithm. APRIORI-KC uses geographic database schemas as prior knowledge to eliminate all association rules which *reproduce* obvious geographic dependences.

The remainder of the paper is organized as follows: Section 2 introduces geographic database schemas and how dependences can be extracted. Section 3 describes the problem of mining spatial association rules with well known geographic dependences and presents the APRIORI-KC algorithm. Section 4 evaluates the algorithm, Section 5 presents the related works, and Section 6 concludes the paper and gives some directions of future work.

II. Geographic Database Schemas

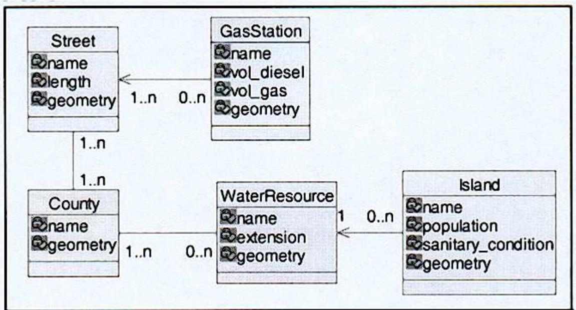
Geographic database schemas are normally extended relational or object-oriented schemas [3]. There is an emerging trend toward extending both Entity Relationship (ER) and Object-Oriented (OO) diagrams with pictograms to provide special treatment to spatial data types [3]. [6] and [7] are approaches which extend ER and OO diagrams for geographic applications. In both ER and OO approaches, relationships among entities are represented through associations with cardinality constraints. In geographic database schemas, these associations may either represent a spatial relationship or a single association, aggregation, etc.

Mandatory associations are represented by cardinality constraints *one-one* and *one-many* [3]-[4]. Geographic data with these association constraints, under rare exceptions, produce well known rules with a 100% confidence.

Fig. 3 shows an example of part of a conceptual geographic database schema, represented in a UML class diagram [8], and part of its respective logical schema for relational and OO databases. The schema in Fig. 3 represents part of the data shown in Fig. 2. Notice that there are many mandatory associations (e.g. gas station and street, street and county, water resource and county, and island and water resource). These dependences, explicitly represented in the schema, produce well known patterns when considered in spatial association rule mining (see in Fig. 2(left) that every gas station intersects one or more streets).

Associations which are not represented as well known dependences in the schema (e.g. gas stations and water resources) may not produce well known patterns. Observe in Fig. 2(right) that there is no well known pattern between gas stations and water resources.

In the logical level, mandatory relationships expressed by cardinalities *one-one* and *one-many* normally result in foreign-keys in relational geographic databases, and in pointers to classes, in object-oriented geographic databases [3]-[4].



Part of Relational Schema

```
Create Table Street
(street_code integer,
street_name varchar(30),
street_length varchar(30),
geometry line,
Primary Key (street_code))

Create Table GasStation
(gas_station_code integer,
gas_station_name varchar(30),
gas_station_vol_diesel float,
gas_station_vol_gas float,
geometry point,
  Primary Key (gas_station_code)
  Foreign Key (street_code) References Street)

Create Table Island
(island_code integer,
 island_name varchar(30),
 island_population integer,
 island_sanitary_condition varchar (10),
 geometry point,
  Primary Key (island_code)
  Foreign Key (water_resource_code) References WaterResource)
```

Part of OO Schema

```
public class GasStation {
  private varchar(30) name;
  private  float vol_diesel;
  private float vol_gas;
  private point geometry;
  Street theStreet;

  public GasStation() {  }
}

public class Island  {
  private varchar(30) name;
  private integer population;
  private varchar(10) sanitary_condition;
  private point geometry;
  WaterResource theWaterResource;

  public Island(){  }
}
```

Fig 3. Conceptual and logical geographic database schema

Well known geographic dependences can be either specified by the user or automatically retrieved with processes of reverse engineering [9] if the schema is not available. Many different approaches to extract dependences from relational databases using reverse engineering are available in the literature. For data mining and knowledge discovery in non-geographic databases reverse engineering has been used to understand the data model [10] in legacy systems, or to automatically extract SQL queries [11], but not as prior knowledge to reduce well known patterns.

When provided by the user, a larger set of dependences can be specified; not only associations explicitly represented in the schema, but other geographic domain dependences which produce well known patterns.

Fig. 4 shows an example of a data preprocessing algorithm to extract mandatory *one-one* and *one-many* associations from geographic database schemas. If the database is relational, then the algorithm searches for all foreign keys. For each foreign key, the name of the table which it references is retrieved, as well as the name of the table where the foreign key is specified. The name of both relations is stored in a set of knowledge constraints ϕ. If the database is object-oriented, then the same steps are performed, but searching for classes with attributes which refer to other classes.

```
Given: a relational database schema
Find all foreign_keys
For each foreign_key
  Insert into ϕ the name of the table
  which the foreign_key references and
  the name of the table to which the
  foreign_key belongs

Given: an OO database schema
Find all classes
For each class in the database schema
  If there are references to classes
      Insert the class name and the
      referenced class into ϕ
```

Fig. 4. Algorithm to extract mandatory relationships

In order to evaluate the amount of mandatory associations in real geographic database schemas we analyzed the object-oriented geographic database schema developed by the Brazilian Army. As the terrain model has a large number of object types, common to different schemas, the database schema developed by the Brazilian Army includes most geographic objects abstracted from the real world as well as their associations.

On account of the large number of entities and relationships to be represented, geographic data conceptual schemas are usually designed in different layers of information. The geographic database schema developed by the Brazilian Army is composed of 8 layers (subschemas): edification, infra-structure, hydrography, vegetation, administrative regions, referential, relief, and toponymy. The layer infra-structure, for example, is divided in six sub-schemas, including information about transportation, energy, economy, communication, etc. The hydrography layer, for example, represents geographic objects such as rivers, oceans, lakes, etc.

Information of different layers may be extracted for data mining, and the number of *one-one* and *one-many* relationships varies from layer to layer. For example, the hydrography layer, which is shown in Fig. 5, has a total of 24 geographic objects (16 from its own layer and 8 from other layers) which share 2 *one-many* relationships and 16 *one-one* relationships if super classes are concrete, and more that 20 if super classes are abstract.

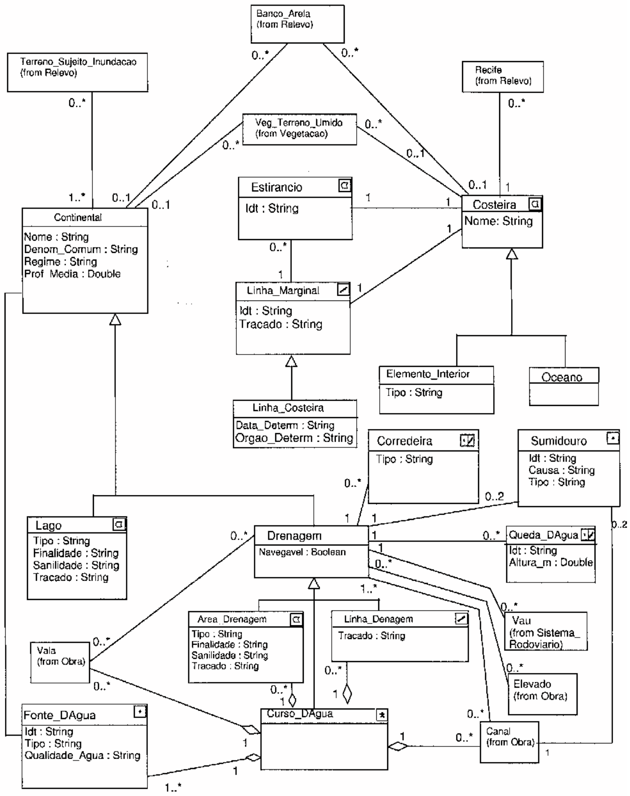


Fig. 5. Conceptual object oriented schema of the Brazilian Geographic Territory (MCOO of EBG - Brazilian Army – STI – DSG - 1°DL)

The infra-structure level, for example, not shown because of space limitations, has 73 geographic objects in its own layer and has relationships with 88 objects in other layers. Among the 88 relationships, 70 are mandatory *one-one* dependences.

The analysis showed that a large number of mandatory well known geographic dependences are explicitly specified in the schema, and if used as prior knowledge to avoid their extraction in association rule mining, a large amount of irrelevant patterns would be eliminated.

III. Mining Spatial Association Rules With Knowledge Constraints

We illustrate the problem of mining spatial association rules without removing explicit geographic domain dependences through an example. Considering a set of elements $\Psi = \{A,B,C,D\}$, all possible combinations of these elements produce the sets: $\{AB\}$, $\{AC\}$, $\{AD\}$, $\{BC\}$, $\{BD\}$, $\{CD\}$, $\{ABC\}$, $\{ABD\}$, $\{ACD\}$, $\{BCD\}$, and $\{ABCD\}$. Without considering any threshold, the number of possible subsets is 11, and the maximum number of rules produced with these subsets is 50, as shown in Table I.

Now consider that the elements C and D have a mandatory association. Notice that there are four subsets in which C and D appear together ({CD}, {ACD}, {BCD} and {$ABCD$}). These four subsets will produce 28 rules, and in every rule, C and D will appear. The result is that 56% of the whole amount of rules is created with the dependence between C and D.

TABLE I
MAXIMUM NUMBER OF COMBINATION SETS AND MAXIMUM NUMBER OF ASSOCIATION RULES

Sets	*Possible Rules*	*Number of Rules*
{AB}	*A→B; B→A*	*2*
{AC}	*A→C; C→A*	*2*
{AD}	*A→D; D→A*	*2*
{BC}	*B→C; C→B*	*2*
{BD}	*B→D; D→B*	*2*
{CD}	*C→D; D→C*	*2*
{ABC}	*A→BC; B→AC; C→AB; BC→A; AC→B; AB→C*	*6*
{ABD}	*A→BD; B→AD; D→AB; BD→A; AD→B; AB→D*	*6*
{ACD}	*A→DC; D→AC; C→AD; DC→A; AC→D; AD→C*	*6*
{BCD}	*D→BC; B→DC; C→DB; BC→D; DC→B; DB→C*	*6*
{ABCD}	*A→BCD; B→ACD; C→ABD; D→ABC; AB→CD; AC→BD; AD→BC; BC→AD; BD→AC; CD→AB; BCD→A; ACD→B; ABD→C; ABC→D;*	*14*

It is important to observe that we cannot just remove C and D from Ψ, because either C or D may have an interesting association with A or B. However, we can avoid the combination of C and D in the same set. This eliminates the possibility of generating rules including both C and D.

In the next sections we describe the formal problem of mining association rules and how APRIORI-KC removes combinations of dependent objects.

A. *Spatial Association Rules*

An association rule consists of an implication of the form $X \rightarrow Y$, where X and Y are sets of items co-occurring in a given tuple [1]. Spatial association rules are defined in terms of spatial predicates, where at least one element in X or Y is a spatial predicate [2]. Spatial predicates represent materialized spatial relationships between geographic elements, such as *close, far, contains, within, touches,* etc. For example, *is_a(x,slum)* $\wedge$ *far_from(x,water_network)* $\rightarrow$ *disease(hepatitis)* (70%) is a spatial association rule with 70% confidence. In [12] we presented an intelligent framework to automatically extract spatial predicates from large geographic databases.

The formal problem statement for defining association rules can be specified as follows: Let $F = \{f_1, f_2, \ldots, f_k, \ldots, f_n\}$ be a set of non-spatial attributes and spatial objects. Let Ψ (dataset) be a set of reference objects T, where each T is a set of predicates (tuple) such that $T \subseteq F$. Each T is represented as a binary vector, with an element $t[k] = 1$, if T contains the attribute f_k, and $t[k] = 0$, otherwise. There is exactly one tuple in the dataset to be mined for each reference object. Considering X as a subset of F, T contains X if, for all f_k in X, $t[k] = 1$. Similarly, being Y a subset of F, T contains Y if, for all f_k in Y, $t[k] = 1$.

In a rule $X \rightarrow Y$, $X \subset F$, $Y \subset F$ and $X \cap Y = \emptyset$. The support s of a predicate set X is the number of tuples in which the predicate set X occurs as a subset. The support of the rule $X \rightarrow Y$ is given as $s(X \cup Y)$.

The rule $X \rightarrow Y$ is satisfied in Ψ with confidence factor $0 \le c \le 1$, if at least c% of the instances in Ψ that satisfy X also satisfy Y. The notation $X \rightarrow Y \mid c$ specifies that the rule $X \rightarrow Y$ has confidence factor of c. More precisely, the confidence factor is given as $s(X \cup Y)/s(X)$.

The problem of mining association rules can be decomposed in two steps: *find all large sets of predicates* - a set of predicates is large if the support is above a certain threshold, and *generate high confidence rules* - support is higher than the minimum support and the confidence is higher than a certain threshold.

Assertion 1. If a predicate set Z is large, then every subset of Z will also be large. If the set Z is not large, then every set that contains Z is not large too. All rules derived from Z satisfy the support constraints if Z satisfies the support constraints.

Considering Assertion1, we propose a third class of constraints, called *knowledge constraints* (ϕ). These constraints will be used to avoid the generation of sets which contain the pairs of dependences specified in ϕ.

B. *The APRIORI-KC Algorithm*

The APRIORI-KC algorithm, shown in Fig. 6, is based on Apriori [13], which has been the basis for dozens of association rule mining algorithms which generate candidate sets, closed sets [14], free sets [15] or any other type of frequent sets.

```
Given: φ, Ψ, minsup
L1 = {large 1-predicate sets};
For ( k = 2; Lk-1 != ∅; k++ ) do  begin
    Ck = apriori_gen(Lk-1); // Generates new
                            // candidates
    Forall T ∈ Ψ do begin
       Ct = subset(Ck, T); // Candidates in t
       forall candidates c ∈ Ct do
            c.count++;
    End;
    Lk = {c ∈ Ck | c.count ≥ minsup};

    If k = 2  // in the second pass
        L2 = L2 - φ ; //removes pairs with
                    // dependences

End;
Answer = ∪k Lk
```

Fig. 6. Pseudo-code of APRIORI-KC to generate large predicate sets without well known dependences

APRIORI-KC removes from the candidate sets all pairs of elements which have geographic dependences. As in Apriori, APRIORI-KC performs multiple passes over the dataset. In the first pass, the support of the individual elements is computed to determine 1-predicate sets. In the subsequent passes, given k as the number of the current pass, the large sets L_{k-1} in the previous pass (k -1) are grouped into sets C_k with k elements, which are called *candidate sets*.

The support of each candidate set is computed, and if it is equal or higher than minimum support, then this set is considered large. This process continues until the number of large sets is zero.

Similarly to [16], which eliminates in the second pass candidate sets that contain both parent and child specified in concept hierarchies, we propose a method to eliminate all candidate sets which contain geographic dependences, independently of any concept hierarchy.

The dependences are eliminated in an efficient way, in one step, in the second pass, when generating candidates with 2 elements. Being ϕ a set of pairs of geographic objects with dependences, which can be extracted from the database schema or provided by the user, when k is 2, all pairs of elements with a dependence in ϕ are removed from C_2.

According to Assertion1, this step *warrants* that the pairs of geographic objects in ϕ will neither appear together in the frequent sets nor in the spatial association rules. This makes our approach effective and independent of any threshold such as minimum support, minimum confidence, lift, etc.

APRIORI-KC eliminates each pair of geographic objects in ϕ (e.g. {C,D}), and avoids the generation not only of the main rule C→D but of all derived rules (e.g. D→C, C→AD) which contain the known dependence. This is the characteristic that differs APRIORI-KC from other algorithms [17]-[18] which eliminate only redundant rules.

IV. Evaluation

For the data mining user, more important than the reduction of the computational time to generate spatial association rules is the elimination of well known geographic domain rules, which is the main objective of this paper. However, as the elimination of the set of pairs which have an association is performed only once, in the second pass, the computational time to generate the rules automatically decreases, since less candidate sets will be generated.

Without considering thresholds of minimum support and minimum confidence, we can quantify the maximum number of generated rules $r_{\max}(n,d)$ as a function with a number n of elements f_k in F and the number d of pairs of dependences in ϕ. For $d = 0$, i.e., without dependences, we have

$$r_{\max}(n,0) = \sum_{i=2}^{n} C_n^i \cdot \left(2^i - 2\right) \tag{1}$$

rules, where C denotes all possible combinations of n in sets with size i. For each of the possible candidate sets, 2^i-2 is the maximum number of rules that can be generated.

By removing one pair of dependences, without considering minimum support and minimum confidence, the maximum number of generated rules is

$$r_{\max}(n,1) = \sum_{i=2}^{n-1} \left(C_n^i - C_{n-2}^{n-i}\right) \cdot \left(2^i - 2\right) \tag{2}$$

where the second combination denotes the number of candidate sets in which the pair of dependences to be eliminated appears.

If two pairs of dependences without overlapping are eliminated, then the maximum number of rules to be generated for $n \geq 6$ elements will be

$$r_{\max}(n,2) = \sum_{i=2}^{n-2} \left(C_n^i - 2 \cdot C_{n-2}^{n-i}\right) \cdot \left(2^i - 2\right) + \sum_{i=4}^{n-2} C_{n-4}^{i-4} \cdot \left(2^i - 2\right) \tag{3}$$

Fig. 7 shows the result of experiments with a different number of geographic objects, where zero, one, and two pairs of dependences were eliminated from the dataset. The reference curve corresponds to the rules mined without dependence elimination. Notice that in the three cases, the number of rules grows exponentially, but the elimination of dependences does considerably reduce the number of rules even when the number of elements increases.

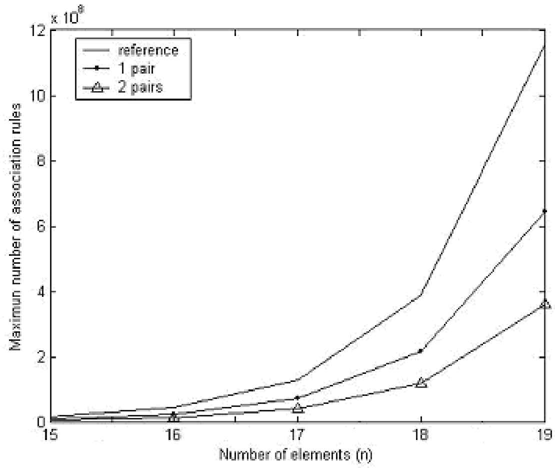


Fig. 7. Maximum number of rules eliminating zero (reference), one, and two pairs of dependences

The elimination of only a few dependences already shows the significance of the method. Notice that the higher the number of well known dependences is, the more significant will the rule reduction be.

The percentage reduction of rules produced when zero, one, and two pairs of dependences are eliminated is shown in Fig. 8. The elimination of one pair generates only 55% of the total number of rules, i.e., eliminates well known rules in 45%. When two pairs are eliminated, only 30% of the rules are created, and the reduction increases to 70%. Notice that even if the number of elements increases, these values represent saturation points for these curves.

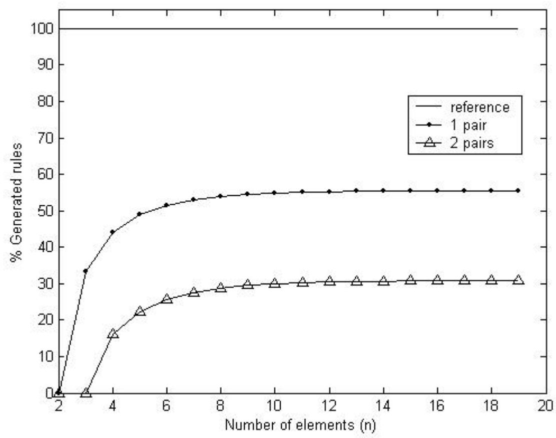


Fig. 8. Percentage reduction of rules considering zero (reference), one, and two pairs of dependences

V. Related Works

Existing approaches for mining spatial association rules do not make use of prior knowledge to reduce the number of well known patterns. Koperski [2] presented a top-down, progressive refinement method. In a first step spatial approximations are calculated, and in a second step, more precise spatial relationships are computed to the result of the first step. Minimum support is used in data preprocessing to extract only frequent spatial relationships. A similar method has been proposed by [19] for mining association rules among geographic objects with broad boundaries. [20] applied Apriori [13] to geographic data at different granularity levels.

In previous work [5] we presented a data preprocessing method using prior knowledge to reduce geographic dependences between the reference object and the relevant objects. However, geographic dependences among relevant objects can only be completely eliminated during the data mining step, as proposed in this paper with APRIORI-KC.

In geographic databases, minimum support can eliminate information which may lead to novel knowledge, while geography domain associations may still remain among the resultant set of rules.

While approaches for non-geographic define different ways to reduce the number of rules [16]-[17]-[18] and investigate the most appropriate threshold [21] or the interestingness [22] of the rules, our approach eliminates the exact dependence which produces non-novel rules, independently of any threshold. Our method avoids the generation of rules known a priori as non-interesting.

VI. CONCLUSIONS AND FUTURE WORK

In this paper we presented a method for mining spatial association rules using prior geographic domain knowledge. Domain knowledge refers to mandatory geographic dependences explicitly represented in geographic database schemas or which are well known by the user. We showed that explicit mandatory relationships produce irrelevant patterns, while the implicit spatial relationships may lead to more interesting rules.

Considering geographic domain dependences as prior knowledge, we proposed a frequent set pruning method that significantly reduces the number of spatial association rules. Besides pruning a large number of rules, all associations that would be created with well known geographic domain dependences are eliminated. Our method eliminates all dependences in one single step, before creating the rules. The result is that more interesting rules will be generated, independently of values of minimum support or confidence.

Experiments showed that independent of the number of elements, one dependence is enough to prune almost half of the total number of spatial association rules, and the higher the number of the dependences, the larger is the reduction.

The main contribution of our approach is for the data mining user, which does not have to analyze hundreds or thousands of rules without novel knowledge.

Traditional association rule mining algorithms that generate *frequent* sets, closed *frequent* sets, or free *frequent* sets eliminate redundant and non-interesting rules. They may significantly reduce the total number of association rules if applied to the geographic domain and the frequent sets were generated with APRIORI-KC.

As future work we will evaluate the problem of mining spatial association rules with knowledge constraints when pairs of geographic dependences overlap. Furthermore we will evaluate the rule reduction and the specification of dependences represented by *n-ary* associations.

VII. ACKNOWLEDGMENT

Our thanks for both CAPES and CNPQ which partially provided the financial support for this research. For the Brazilian Army, which provided the geographic database schema, and for Procempa, for the geographic database.

VIII. REFERENCES

1. R. Agrawal, T. Imielinski, and A. Swami, “Mining association rules between sets of items in large databases,” In Proceedings of ACM-SIGMOD International Conference Management of Data, pp. 207-216, 1993.
2. K. Koperski, and J. Han, “Discovery of spatial association rules in geographic information databases,” In Proceedings of the SSD 4th international Symposium in large Spatial Databases, (SSD'95) Springer, Portland, Maine, USA, pp. 47-66, 1995.
3. S. Shekhar, and S. Chawla, *Spatial databases: a tour,* Prentice Hall, Upper Saddle River, NJ, 2003.
4. R. Elmasri, and S. Navathe, *Fundamentals of database systems*, 4th ed., Addison Wesley, 2003.
5. V. Bogorny, P.M. Engel, and L.O. Alvares, “GEOARM: an Interoperable Framework to Improve Geographic Data Preprocessing and Spatial Association Rule Mining”. In Proceedings of the 18th International Conference on Software Engineering and Knowledge Engineering, (SEKE'06). Jul. 2006, San Francisco, California, 2006 (accepted for publication) .
6. C. Parent, S. Spaccapietra, E. Zimanyi, P. Donini, C. Plazanet, and C. Vangenot, “Modeling spatial data in the MADS conceptual model,” In Proceedings of the 8th International Symposium on Spatial Data Handling, (SDH'98), Vancouver, pp. 138-150, 1998.
7. K.A.V. Borges, A. H. F. Laender, C. Davis JR. “OMT-G: an object-oriented data model for geographic applications. *GeoInformatica,* vol. 5(3), 2001, pp.221-260.
8. G. Booch, J. Rumbaugh, and I. Jacobson, *The unified modeling language: user guide*, Addison-Wesley, 1998.
9. E.J. Chifosky, and J. H. Cross, “Reverse engineering and design recovery: a taxonomy”. *IEEE Software*, Jan .1990.
10. S. MCKearney, and H. Roberts, “Reverse engineering databases for knowledge discovery”. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD’96). Portland, Oregon, pp.375-378, 1996.
11. P. Shoval, and N. Shreiber, “Database reverse engineering: from the relational to the binary relationship model”. Data and Knowledge Engineering. vol. 10, 1993, pp. 293-315.
12. V. Bogorny, P.M. Engel, L.O. Alvares, “A Reuse-Based Spatial Data Preparation Framework for Data Mining,” In Proceedings of the 17 th International Conference on Software Engineering and Knowledge Engineering, (SEKE'05), Taiwan, China, pp. 649-652, 2005.
13. R. Agrawal, and R. Srikant, “Fast algorithms for mining association rules. In Proceedings of the 20th Intermatopmaç Conference on Very Large Databases (VLDB’94). Santiago, Chile, 1994.
14. M. Zaki, and H. Ching-Jui, “CHARM: An efficient algorithm for closed itemset mining”, Second SIAM International Conference on Data Mining, Arlington, pp. 457-473, 2002.
15. J.F. Boulicaut, M. Jeudy, “Mining free itemsets under constraints”, In Proceedings of the International Database Engineering & Applications Symposium, (IDEAS '01), *IEEE Computer Society,* Washington, DC, USA, pp. 322-329, 2001.
16. Srikant, R., Agrawal, R. “Mining generalized association rules”, In Proceedings of the 21st International Conference on Very Large Databases (VLDB’1995), Zurich, Switzerland, pp. 407-419, 1995.
17. Zaki, M. “Generating non-redundant association rules”, In Proceedings of the 6th ACM (SIGKDD’2000) International Conference on Knowledge Discovery and Data Mining, Boston, Massachussets, USA, pp.34-43, 2000.
18. B. Liu, W. Hsu, and Y. Ma, “Pruning and summarizing the discovered associations”. Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining. San Diego, California, pp. 125–134, 1999.
19. E. Clementini, P. Di Felice, K. Koperski, “Mining multiple-level spatial association rules for objects with a broad boundary”. *Data & Knowledge Engineering*. vol. 34(3), 2000, pp.251-270.
20. J. Mennis, J. Liu, “Mining association rules in spatio-temporal data: an analysis of urban socioeconomic and land cover change.” *Transactions in GIS* 9 (1), 2005, pp. 5-17.
21. P.N. Tan, V. Kumar, and J. Srivastava, “Selecting the right interestingness measure for association patterns”. In Proceedings of The eighth International Conference on Knowledge Discovery and Data Mining, ACM Press, pp. 32-41, 2002.
22. A. Silberschatz, and A. Tuzhilin, “What makes patterns interesting in knowledge discovery systems”. Knowledge and Data Engineering, *IEEE Transactions on Knowledge and Data Engineering*. vol. 8(6), pp. 970–974, Dec. 1996.