# QUERYING AND MINING SPATIOTEMPORAL ASSOCIATION RULES

Hana Alouaoui, Sami Yassine Turki
*LTSIRS Laboratory of Remote Sensing and Information Systems with Spatial Reference*
*ENIT National Engineering School of Tunis, Tunis El Manar University, Tunis, Tunisia*

Sami Faiz
*LTSIRS Laboratory of Remote Sensing and Information Systems with Spatial Reference*
*ENIT National Engineering School of Tunis, Tunis El Manar University, Tunis, Tunisia*

Keywords: Spatiotemporal database, Topological relationships, Spatiotemporal data mining, Data mining query language.

Abstract: This paper presents an approach for mining spatiotemporal association rules. The proposed method is based on the computation of neighborhood relationships between geographic objects during a time interval. This kind of information is extracted from spatiotemporal database by the means of special mining queries enriched by time management parameters. The resulting spatiotemporal predicates are then processed by classical data mining tools in order to generate spatiotemporal association rules.

## 1 INTRODUCTION

Extracting interesting and useful patterns from spatial and temporal sets is more difficult than extracting corresponding patterns from traditional data due to the complexity of spatial and temporal data types and spatial relationships changing over the time.

Our contribution is to process the spatiotemporal components by computing neighborhood relationships between geographic objects during a time interval. This step is achieved by data mining queries enriched by time management tools. Then data mining techniques are applied on the resulting spatiotemporal predicates in order to mine spatiotemporal association rules.

In the next section we make an overview of the existing data mining query languages, the section 3 describes our proposed approach of spatiotemporal association rules mining. Finally, in the section 4 we summarize the main conclusions of this paper and point out directions for current and future work.

## 2 STATE OF THE ART: QUERY LANGUAGES AND KNOWLEDGE DISCOVERY IN DATABASES

The high availability of huge databases - rich in hidden information beyond human's ability to retrieve manually- and the prominent necessity of information and knowledge extraction from such data, have demanded valuable efforts from the scientific community. Finding tools and techniques aiming to analyze these huge data repositories is a subject dealt by the field of Knowledge Discovery in Databases (KDD) (Fayyad et al., 1996).

There have been a number of contributions dealing with different aspects of this problem by proposing structured languages for KDD specification. These languages follow SQL patterns and provide techniques for data preprocessing such as accessing, cleaning, transforming, deriving and mining data (Boulicaut and Masson, 2005).

These languages can integrate background knowledge, like concept hierarchies and can define thresholds (eg: support, confidence; in the case of


Alouaoui H., Yassine Turki S. and Faiz S..
QUERYING AND MINING SPATIOTEMPORAL ASSOCIATION RULES.
DOI: 10.5220/0003636303940397
In *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval* (KDIR-2011), pages 394-397
ISBN: 978-989-8425-79-9
Copyright © 2011 SCITEPRESS (Science and Technology Publications, Lda.)

association rules extraction) in order to extract just the most interesting patterns (Boulicaut and Masson, 2005). A Data Mining Query Language DMQL has been proposed in (Han et al., 1996) for mining association rules using concept hierarchies (Han, 1995) as background knowledge. However, there is just one practical application of DMQL found in the DBMiner (Dbminer, 2000) where it is used as a task description resource.

Another work was proposed by (Meo et al., 1996) is based on a new operator, named MINE RULE, designed as an extension of the SQL language in order to discover association rules.

Other languages have been built on the principles of relational databases (Imielinski and Virmani, 1999), (Wang and Zaniolo, 2003), (Marcelino et al., 2004). They follow the SQL patterns with resources for accessing, cleaning, transforming, deriving and mining data, beyond knowledge manipulation.

A further important field that needs to get a big attention is the knowledge discovery from spatial databases. Complex data types, intrinsic relations between spatial components and non-spatial components as well as relationships between data themselves make the spatial data mining more difficult. This explains the small number of data mining query languages that have been proposed for spatial data. The GMQL (Gegraphic Mining Query Language) proposed by (Han et al., 1997) is an extension of DMQL to support spatial data mining.

Another approach based on the transformation of a spatial database into an inductive database was proposed by (Malerba et al., 2004). The proposed language needs a complex data preprocessing tasks in order to formulate the queries.

A spatiotemporal data mining query language was proposed in (Chen and Zaniolo, 2000). The $SQL^{ST}$ sees reality as instantaneous sequences of moving objects (Manco et al., 2008), (Bogorny et al., 2008), (Erlend and Mads, 2010) and is limited to mine knowledge from trajectories evolving in space and time. This language is built on the basis of a temporal data mining query language proposed by (Chen and Petrounias, 1998).

All of these languages dealt with traditional, temporal, or spatial data. They treated separately the space and the time. The proposed languages merging space and time aspects were simply limited to the trajectory of moving objects. To the best of our knowledge no data mining query language has been proposed in order to cope with the discrete evolution of spatial data over the time. Our problematic is to mine knowledge from discrete evolving objects like parcels or river changing of shape during large time intervals. Our proposed queries are settled on a combination of GMQL and time features.

# 3 THE PROPOSED APPROACH FOR MINING SPATIOTEMPORAL ASSOCIATION RULES

Our approach aims to extract spatiotemporal association rules. In order to achieve this objective two phases should be accomplished; processing the spatiotemporal components and applying data mining techniques. Processing the spatiotemporal components phase is focused on the enrichment of the spatial association rules extraction with time concept integration. This means that the mining of spatial predicates (spatial relationships between geographic entities such as: close, far, contains, within....) will be done during time intervals. In order to accomplish this phase special queries merging data mining and time management tools were applied. The extracted spatiotemporal predicates will be considered as the item-sets treated by the data mining algorithm Apriori (Agrawal and Srikant, 1994) in order to mine spatiotemporal association rules.

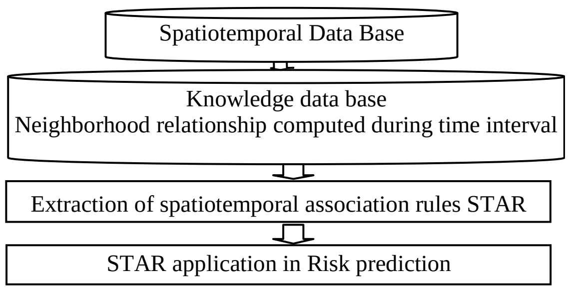


Figure 1: The Architecture of the proposed approach.

## 3.1 Spatial Association Rules Extraction

The spatial association rules are extensions of association rules with spatial features. Associations highlighted include the proprieties of neighbor objects and their neighborhood relationship (Zeitouni and Yeh, 1999). The generated rules have the form: X→Y where X and Y are sets of spatial and non spatial predicates. Thus, the spatial association rules have the following form (1):

$$P1 \cap \ldots \cap Pm \rightarrow Q1 \ldots \cap Qn \qquad S(\%) \quad C(\%) \qquad (1)$$

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

Where at least one of Pi or Qi is spatial. A rule is always provided with two measures: support (s %) and confidence (c %). The support is percentage of transactions that satisfy X and Y among all transactions from the transaction base.

The Confidence is the percentage of transactions that verify the conclusion of a rule among those that satisfy the premise data (Turki and Faïz, 2009).

The determination of the SAR goes through 3 phases (Bogorny, 2006):

- Calculation of spatial predicates (spatial relationships between geographic entities). (Han et al., 1997); (Koperski, 1999); (Koperski and Han, 1995).
- Generation of frequent itemsets: an itemset is frequent if at least its support is equal to a minimum threshold (minsup).
- The extraction of spatial association rules.

We noted that the step 2 and 3 of the process of data mining have received great importance in the literature and were considered as the major problems and were designated by frequent pattern mining and association rule mining. The first step of spatial relationships calculation is the largest typically because the effectiveness and efficiency of the extracted rules is based on these relationships. These spatial relations are the main characteristic of spatial data and must be taken into account in the process of knowledge extraction and this is the primary characteristic that distinguishes spatial data mining and classic data mining (Koperski, 1999); (Bogorny, 2006).

## 3.2 The Proposed Query

We have 3 Classes (tables); Road, Town and Oued (which means a kind of rivers existing in the north of Africa).

The Town is our target object or (reference object), the road and the river are its neighbors or relevant task objects.

The time is stored as an attribute in the table and we adopted the notion of valid time.

The information in the data base is up-to-date with the technique of attribute versioning; the new values or states of an object are stored with the new interval of validity. The relationships between the objects are defined on the basis of distance parameters given by domain experts (e.g: Besides (Distance > 0m and <= 50m))

Example of the proposed query:

We have two time intervals I1 and I2; I1 ([01-01-2000, 31-12-2004]) and I2 ([01-01-2005, 31-12-2009])

```
Q1: Mine spatio-temporal associations
describing Town with respect to
Topology (T.geo, O.geo),
Topology (T.geo, RD.geo),
Tvalidtime_, Ovalidtime, Rdvalidtime
From Town T, Oued O, Road RD
Where distance (T.geo, O.geo) <= "2km"
AND distance (T.geo, RD.geo) <="500 m"
VALID IN [01-01-2000, 31-12-2004]
```

The results of these queries are collected and organized in a knowledge base containing information related to dependencies between spatial objects computed during time intervals.

Other neighborhood relationships are computed and describe some prohibited relations between the spatial objects.

Example: the relation (oued_contains) is a vulnerable situation because if an oued (river) covers an urban zone may possibly cause an inundation.

## 3.3 Mining Spatiotemporal Association Rules

The computed spatiotemporal predicates were used in the mining of spatiotemporal association rules leading to decide about the possibility of natural risk occurrence.

A set of STAR (Spatiotemporal Association Rules) was generated.

For example, the following association rule is derived from the spatiotemporal data set.

oued_touches_I1=yes oued_crosses_I2=yes ==> oued_contains_I2=yes (40%, 90%)

This rule shows the temporal evolution of the river (oued) that was near of the town ((oued_touches), (distance > 50m and <=200m)) at the time interval (I1) then it crossed it (oued_crosses) at (I2). As a result of this shape change a prohibited relationship ((oued_contains), (distance < 0m)) appeared which can be explained by an inundation risk occurrence.

The values 40% and 90% indicate respectively the support and the confidence of the rule.

The most meaningful rules will be stored as learning examples and will be processed by a learning system (e.g. neural network) in order to identify unknown future risks. This will be the object of our future work.

# 4 CONCLUSIONS AND FUTURE WORK

In this paper we highlight the necessity to incorporate the temporal measures in studying the evolution of geographical objects over the time.

In order to achieve this objective we proposed an approach aiming to extract knowledge from a spatiotemporal database by the means of spatiotemporal mining queries merging both data mining and time management concepts.

As future work, the most meaningful rules will be processed by a learning system (e.g. neural network) in order to identify unknown future risks.

We will also evaluate the possibility to implement our approach in a wider range of risk prediction applications including appropriate input parameters suitable for the study region and the kind of risk to predict.

# REFERENCES


Agrawal, R., Srikant, R. 1994. Fast algorithms for mining association rules in large databases. Research Report RJ 9839, *IBM Almaden Research Center*, San Jose, California.

Bogorny, V., 2006. Enhancing spatial association rule mining in geographical database. Thesis presented In partial fulfillment of the requirements for the degree of *doctor in computer science, federal university of*" *Rio Grande du sul*", Porto Alegre.

Bogorny, V., Bart, K., Luis, O., 2008. A Spatio-temporal Data Mining Query Language for Moving Object Trajectories. Technical Report TR-357. *Federal University of Rio Grande do Sul*, Porto Alegre, Brazil.

Boulicaut, J. F. and Masson, C., 2005, Data Mining Query Languages. In The Data Mining and Knowledge Discovery Handbook, O. Maimon and L. Rokach (Eds) (Springer), pp. 715{727.

Chen, X., Zaniolo, C., 2000. SQL$^{ST}$: A Spatiotemporal Data Model and Query Language. *International Conference on Conceptual Modeling*.

Chen, X., Petrounias, I., 1998, Language Support for Temporal Data Mining. In *Proceedings of the Proceedings of the Second European Symposium on Principles of Data Mining and Knowledge Discovery* (London, UK: Springer-Verlag), pp. 282{290.

Dbminer Technology 2000.Inc. DBMiner Enterprise 2.0. Available at DBMiner Technology site. URL: http:// www.dbminer.com.

Erlend, T., Mads, N., 2010. Representing topological relationships for spatiotemporal objects. In GEOINFORMATICA. (published with open access at Springerlink.com).

Fayyad, U., Piatesky-Shapiro, G., Smyth, P., 1996. From Data Mining to Knowledge Discovery: An Overview. Advances in KDD and Data Mining, AAAI,

Han, J., Koperski, K., Stefanovic, N., 1997. GeoMiner: A System Prototype for Spatial Data Mining. *In ACM-SIGMOD Int'l Conf. On Management of Data* (SIGMOD'97), Tucson, Arizona.

Han, J., Fu, Y., Wang, W., Koperski, K. and Zaiane, O., 1996, Dmql: A data mining query language for relational databases. *In Proceedings of the SIGMOD'96 Workshop on Research Issues in DataMining and Knowledge Discovery*, Montreal, Canada, pp. 27{33.

Han, J., 1995, Mining Knowledge at Multiple Concept Levels. *In Proceedings of the CIKM* (ACM), pp. 19{24.

Imielinski, T. and Virmani, A., 1999, MSQL: A Query Language for Database Mining. Data Mining and Knowledge Discovery, 3, 373{408.

Koperski, K., 1999. A progressive refinement approach to spatial data mining. Doctorate Thesis. *In Simon Fraser University*.

Koperski, K., Han, J., 1995. Discovery of spatial association rules in Geographic Information Databases. *In proc.4th Int'Symp on large Databases* (SSD'95), pp47-66, Portland.

Manco, G., Baglioni, M., Giannotti, F., Kuijpers, B., Raffaet, A., Renso, C., 2008. Querying and Reasoning for Spatiotemporal Data Mining F. Giannotti and D. Pedreschi (eds.) *Mobility, Data Mining and Privacy*.c Springer-Verlag Berlin Heidelberg .

Marcelino, P., and Robin, J., 2004. SKDQL, a structured language to specify knowledge discovery processes and queries. Lecture Notes in Artificial Intelligence 3171, Springer.

Malerba, D., Appice, A. and Ceci, M., 2004, A Data Mining Query Language for Knowledge Discovery in a Geographical Information System. *In Proceedings of the Database Support for Data Mining Applications*, pp. 95{116.

Meo, R., Psaila, G. and Ceri, S., 1996, A New SQL-like Operator for Mining Association Rules. In *Proceedings of the VLDB*, T.M. Vijayaraman, A.P. Buchmann, C. Mohan and N.L. Sarda (Eds) (Morgan Kaufmann), pp. 122{133

Roshan, N., Asghar, A., 1996. The Management of Spatio-Temporal Data in a National Geographic Information System. Springer.

Turki, Y., Faïz, S., 2009. Apport des règles d'association spatiales pour l'alimentation automatique des bases de données géographiques. *International journal of Geomatic*. Hermès–Lavoisier editions, Paris, France, Vol. 19/ N°1/2009, pp.27-44.

Wang, H. and Zaniolo, C., 2003, ATLaS: A Native Extension of SQL for Data Mining. *In Proceedings of the SDM*, D. Barbar_a and C. Kamath (Eds) (SIAM).

Zeitouni, K., Yeh, L., 1999. Les bases de données spatiales et le data mining spatial. *International journal of Geomatic*, Vol 9, N° 4.


SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS