

Facility management and mining spatio-temporal data

Luboš Popelínský¹ and Petr Glos²

¹Knowledge Discovery Lab, Faculty of Informatics, Masaryk University

²Institute of Computer Science, Masaryk University

Botanická 68a, 60200 Brno

popel@fi.muni.cz, glos@ics.muni.cz

Abstract. We present ongoing project on mining spatio-temporal frequent patterns from facility management data. We introduce data on facility management in MU. After introduction of spatio-temporal first-order patterns we focus on two tasks that are of importance for facility management - mining frequent patterns and mining rare events in spatio-temporal data.

Key words: facility management, data mining, spatio-temporal data, frequent patterns, association rules

1 Facility management

The definition of FM provided by the European Committee for Standardization (CEN) and ratified by BSI [British Standards](#) is: “Facilities management is the integration of processes within an organization to maintain and develop the agreed services which support and improve the effectiveness of its primary activities”. A definition provided by the [International Facility Management Association](#) (IFMA) <http://www.ifma.org/> facility management is a profession that encompasses multiple disciplines to ensure functionality of the built environment by integrating people, place, process and technology. Masaryk University maintains a digital version of building passport that currently consists of approximately 200 buildings and 17,000 rooms. For their representation in a geodatabase, several distinct constructions (building primitives) were defined and every building is made up of these objects. The resulting passport data is available to university employees, students and even for public via the internet/intranet as well as it is used by other university's information systems. The university plans to create also a technological passport, which is closely related to the building passport. Additionally, the building passport is used to generate 3D models of the buildings. Masaryk University also implements Building Management System (BMS) based on BACNet open protocol. BMS provides a means of storing a historical building operation data in the relational database. BMS database contains data of building environment (e.g. room temperature, room humidity, room air pressure difference), status of technologies (e.g. run/stop, fault) and consumptions (electric energy, cold and hot water, gas). The common goal of facility management methods is increasing of building operation cost. We need to recognize dependencies and relationships between BMS data to find out where and how we can cut down operation cost. It is not surprising that a need for deep analysis of this data arised. Besides of visual analytics tools, and in collaboration with them,

we aim at applying data mining methods - frequent patterns mining and association rules mining [5, 6].

2 Mining frequent patterns

Frequent patterns (also called large itemsets) have been originally defined as propositional formulas that are true for at least a given fraction of items in a database [1]. This fraction is called a minimal support. The example is a set of baskets of consumers in a supermarket. In this case, the frequent patterns brings an information about products that appear frequently together in those baskets.

A frequent pattern in predicate logic is a logical conjunction of elementary formulas (atoms) that is frequent for a given data. Here we focus on spatio-temporal logic which extends a predicate logic with temporal operators (e.g. AFTER, BEFORE, ALWAYS, SOMETIMES) and spatial functions (e.g. LEFT-TO, INCLUDED, SOUTH-OF). An example (from windstorm data analysis) is a formula „AFTER a wind K, in the period 1971-72 ALWAYS a wind was strong“ .

3 Spatiotemporal frequent patterns

For this work, spatiotemporal data are supposed to be a sequence of events. An *event* has a unique identifier and is connected with an explicit time instant. In the case that the data not contain an explicit time attribute, it can be substituted with an order of this event in the sequence. At least one attribute must be spatial. It can be x- and y-coordinates (e.g. in windstorm data coordinates of a place) or identifier of an area (e.g. the name of a district). There is no limit on the number of events with the equal time stamp. We also allow attributes of complex type: not only atomic but also of the type of list [9].

A *domain knowledge* is a set of predicate definitions. A *spatiotemporal pattern* (or shortly *pattern*) is a conjunction of non-spatiotemporal and spatiotemporal atoms. Negation is not allowed in a pattern. A *non-spatiotemporal atom* is either of the form *Attribute Operator Value* where an operator is '=' for categorical attributes and '=', '<=', '<' for numerical attribute, or is defined by a predicate from domain knowledge that does not have a temporal attribute as its argument. A *spatiotemporal atom* can be temporal – NEXT, ALWAYS, SOMETIMES – or spatial , e.g.. *dc(X,Y)* (X is disconnected from Y).

The problem of mining spatiotemporal maximal frequent patterns is then to find all frequent spatiotemporal patterns, i.e. those that cover at least M examples (M is usually called a *minimal support* that cannot be further refined without decreasing support below M).

In [9] we introduced a new version of refinement specialization operator for efficient mining in spatiotemporal data. It has been implemented in ILP system RAP. RAP [3] is a tool for mining first-order maximal frequent patterns that employs different search strategies for mining long patterns. Frequent patterns learned with

RAP has been successfully used as new features for knowledge discovery in mining medical data (STULONG), in information extraction from biomedical text and as well as for classification of small pieces of text in reports on flood.

4 Spatial co-location patterns

A co-location pattern is a group of spatial features/events that are frequently co-located in the same region [8]. More formally, a set of spatial features form a pattern if, for each spatial feature, at least $s\%$ instances of that feature form a clique with some instance of all the rest features in the pattern for a given neighborhood relationship. The parameter s is called the participation index. A neighborhood relation can be defined as a distance (e.g. Euclidean), as a topological relation or something else (see [7] for various neighborhood relations).

Mining spatial co-location patterns differs from frequent pattern mining. Instead of item set and a minimal support in frequent pattern mining we have spatial feature set and spatial interestingness measure.

5 Mining rare events

In facility management data mining it is also important to find patterns that are rare but important for precaution. Such event is e.g. fire (or an increase of a temperature), fast repeated switching on/off of a device, or water pipe disruption. In this case it is not a frequent pattern what we are looking for but rather a frequent correlation – mostly spatial or temporal – of two or more attributes. In [8] a novel method based on Apriori [1] algorithm has been described. They introduce a new measure, *maximal participation ratio*, which allows finding spatial co-location patterns in the presence of rare spatial features.

6 Finding spatio-temporal patterns in facility management data

We will adapt two approaches described above to facility management data. We will focus both to mining frequent patterns and to mining rare events.

In the case of frequent patterns mining the first goal is choose/develop appropriate spatiotemporal logic that will be a refinement of the general case introduced in [9]. It contains a definition of neighborhood relations (and consequently spatial relations) that are the most appropriate for facility management data, general enough and efficient to evaluate. Multirelational data mining methods [4] seems be the most convenient because they, first, allow to use domain knowledge in a natural way, and, second, can be easily incorporated with e.g. constraint logic programming.

We will also explore how the existing methods for mining spatiotemporal patterns [3] are related to, or can be modified for, mining co-location patterns. We will look

also for search strategies and post-processing methods that allow adapting existing algorithms for finding rare events in efficient way.

We will extend the concept of maximal participation index introduced in [8] for learning in more powerful spatiotemporal logic.

Following [2] we will look for pattern and rule measures that enable to limit search space and filter the most interesting patterns.

References

1. Agrawal R., Srikant R. Fast Algorithms for Mining Association Rules. In *Proc. 20th Int. Conf. Very Large Data Bases, VLDB 1994*, pp. 487-499.
2. Azevedo, P. and Jorge A. M. (2007) "Comparing Rule Measures for Predictive Association Rules", in *Proceedings of ECML'07* pp 510-517.
3. Blat'ák J., Popelínský, L.: Mining first-order maximal frequent patterns. *Neural Network World* 5, 4, pp. 381-390
4. Džeroski S., Lavrač N. *Relational data mining*. Springer Verlag 2001.
5. Glos, P. Building and Technology Passport of Masaryk University. In *Proceedings of 2008 ESRI International User Conference*. San Diego, California : ESRI Press, 2008
6. Glos, P. Using ArcGIS for Visualizing Historical Data from BMS. In *Proceedings of 2009 ESRI International User Conference*. San Diego California : ESRI Press, 2009
7. Ester M. et al. Spatial Data Mining: A Database Approach. *Advances in spatial databases: 5th international symposium, SSD '97*, Berlin, 1997.
8. Huang Y., Pei J., Xiong H. Mining C-Location Patterns with Rare Events from Spatial Data Sets. *Geoinformatika* 10, 2006, pp. 239-260.
9. L. Popelínský, J. Blat'ák: Toward mining of spatiotemporal maximal frequent patterns. In *Proceedings of ECML/PKDD Workshop on Mining Spatio-Temporal Data (MSTD)*, Porto 2005.
10. Simoff S. J., Böhlen, M. H., Mazeika A. (eds.) (2008) *Visual Data Mining*. LNCS 4404 Springer Verlag 2008.