

Mining spatio-temporal data

**Gennady Andrienko · Donato Malerba ·
Michael May · Maguelonne Teisseire**

© Springer Science + Business Media, LLC 2006

Spatio-temporal data mining is an emerging research area dedicated to the development and application of novel computational techniques for the analysis of large spatio-temporal databases. The main impulse to research in this subfield of data mining comes from the large amount of

- spatial data made available by GIS, CAD, robotics and computer vision applications, computational biology, and mobile computing applications;
- temporal data obtained by registering events (e.g., telecommunication or web traffic data) and monitoring processes and workflows.

Both the temporal and spatial dimensions add substantial complexity to data mining tasks.

First of all, the spatial relations, both metric (such as distance) and non-metric (such as topology, direction, shape, etc.) and the temporal relations (such as before and after) are information bearing and therefore need to be considered in the data mining methods.

Secondly, some spatial and temporal relations are implicitly defined, that is, they are not explicitly encoded in a database. These relations must be extracted from the data and there is a trade-off between precomputing them before the actual mining process starts (eager approach) and computing them on-the-fly when they are actually needed (lazy approach). Moreover, despite much formalization of space and time relations available in spatio-temporal reasoning, the extraction of spatial/temporal relations implicitly defined in the data introduces some degree of fuzziness that may have a large impact on the results of the data mining process.

G. Andrienko · M. May
Fraunhofer Institut Autonome Intelligente Systeme (FhG AIS), Sankt Augustin, Germany

D. Malerba (✉)
University of Bari, Bari, Italy
e-mail: malerba@di.uniba.it

M. Teisseire
Université Montpellier 2—LIRMM—CNRS, Montpellier, France

Thirdly, working at the level of stored data, that is, geometric representations (points, lines and regions) for spatial data or time stamps for temporal data, is often undesirable. For instance, urban planning researchers are interested in possible relations between two roads, which either cross each other, or run parallel, or can be confluent, independently of the fact that the two roads are represented by one or more tuples of a relational table of “lines” or “regions”. Therefore, complex transformations are required to describe the units of analysis at higher conceptual levels, where human-interpretable properties and relations are expressed.

Fourthly, spatial resolution or temporal granularity can have direct impact on the strength of patterns that can be discovered in the datasets. Interesting patterns are more likely to be discovered at the lowest resolution/granularity level. On the other hand, large support is more likely to exist at higher levels.

Fifthly, many rules of qualitative reasoning on spatial and temporal data (e.g., transitive properties for temporal relations after and before), as well as spatio-temporal ontologies, provide a valuable source of domain independent knowledge that should be taken into account when generating patterns. How to express these rules and how to integrate spatio-temporal reasoning mechanisms in data mining systems are still open problems.

Additional research issues related to spatio-temporal data mining concern visualization of spatio-temporal patterns and phenomena, scalability of the methods, data structures used to represent and efficiently index spatio-temporal data.

For this special issue, we have selected seven papers from 14 candidates, which addresses many of the research issues reported above and present foundations, new concepts and application examples of mining spatio-temporal data. Three of them constitute an updated and extended version of papers presented at the Workshop on Mining Spatio-Temporal Data, Porto (Portugal), 3rd October 2006, chaired by the guest editors of this special issue. The workshop was a satellite event of the 16th European Conference on Machine Learning (ECML’05) and of the 9th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD’05). Due to editorial constraints, only five papers are reported in this issue, while the other two will appear in the next issue.

To facilitate orientation for the reader of this special issue, we summarize the main ideas behind the individual papers in the following.

The first paper, “Spatial associative classification: propositional vs. structural approach” by A. Appice and M. Ceci, introduces an associative approach to the classification of spatial objects and compares two different ways of exploiting association rules, namely, to define features of a propositional classifier or to define rules of a first-order naïve Bayesian classifier. Association rules are described in a first-order logic language and are extracted by means of the SPADA method, which supports the analysis at different levels of granularity, considers auto-correlation between spatial objects, prevents the generation of useless spatial patterns through the specification of an explicit language bias, and exploits some background knowledge to change the representation of spatial patterns, making it more abstract (human-comprehensible) and less tied to the physical representation of spatial objects in the database. The propositional and the structural associative classifiers are tested on two spatial data sets.

In “Mining changing regions from access-constrained snapshots: a cluster-embedded decision tree approach”, I. Pekerskaya, J. Pei and K. Wang present an approach to perform change detection on a given set of spatial data snapshots taken

at various temporal instants. Due to data access constraints, such as privacy concerns and limited online availability, data snapshots are summarized in the form of cluster-embedded decision trees. Change detection involves comparing two cluster-embedded decision trees in order to find those regions that have different class labels. The proposed method is tested on USA Census data.

The problem of mapping spatial relations between homogeneous spatial objects into a single dimension that preserves locality of reference within the data is investigated in the paper by D. Guo and M. Gahegan, entitled “Spatial ordering and encoding for geographic data mining and visualization”. This mapping is useful for the application of general-purpose data mining techniques to spatial data. The authors present nine ordering/encoding methods and design a comprehensive set of measures to evaluate them with synthetic data. They show that the optimal ordering/encoding with the complete-linkage clustering consistently gives the best overall performance, although the primary factor that controls the ordering/encoding quality remains in the clustering method.

The paper “Time-focused clustering of trajectories of moving objects”, by M. Nanni and D. Pedreschi, presents a study on clustering trajectories of mobile objects (e.g., mobile phones). Problems faced in this spatio-temporal data mining task concern the identification of the proper spatial granularity level, the selection of the significant temporal subdomains, the choice of the most promising clustering method, and the formalization of the notion of (dis)similarity among trajectories. The authors recommend a density-based approach to trajectory clustering, and stress the importance of “temporal focusing” to isolate the clusters of higher quality. Both aspects have been tested on a data set automatically generated by a synthesizer of trajectory data.

The paper “Mining sequential patterns from data streams: a centroid approach” by A. Marascu and F. Masegla is devoted to the analysis of data along the temporal dimension. It shows that classical sequential pattern mining methods cannot be used in a data stream environment and introduces an algorithm for the approximate discovery of sequential patterns in data streams. The algorithm is based on the summarization of batches of transactions by means of a sequences alignment method. By storing frequent sequences in a prefix tree structure it is possible to detect frequent sequential patterns with a very low support. The proposed algorithm has been tested on web usage data streams.

The sixth article (forthcoming), is entitled “Knowledge discovery from spatial transactions” and authored by S. Rinzivillo and F. Turini”. It proposes a set of spatial operators to generate “spatial” transactions, each of which include attributes of some reference object and its neighbors. Properties of spatial transactions are exploited in adaptations of two data mining algorithms for decision tree building and association rules extraction.

The last paper (forthcoming), “Towards a new approach for mining maximal frequent itemsets over data stream” by C. Raïssi and P. Poncelet is related to extracting frequent itemsets over data streams. The authors have proposed a technique characterised by a new representation structure of the itemsets and user-based requests of the frequent itemsets over an arbitrarily time interval. Experimental results prove the efficiency of the method with respect to the constraints of data stream processing.

The papers in this special issue constitute only a small sample of recent research related to mining spatio-temporal data. However, they give the reader a sense of

some of the most challenging problems that arise when tackling this kind of data. There is no doubt that this is a very promising research area, which will offer researchers a wide range of topics. We hope that this issue will draw added attention to this field and stimulate further research.

Finally, we want to take this opportunity to thank the reviewers for their critical comments on these manuscripts and the authors for their timely and careful revisions.