Spatio-Temporal Data Mining: From Big Data to Patterns

Maguelonne Teisseire
UMR TETIS (Cirad, Irstea, AgroParisTech, CNRS) – France
maguelonne.teisseire@irstea.fr
Web: www.textmining.biz

Abstract

Technological advances in terms of data acquisition enable to better monitor dynamic phenomena in various domains (areas, fields) including environment. The collected data is more and more complex - spatial, temporal, heterogeneous and multi-scale. Exploiting this data requires new data analysis and knowledge discovery methods. In that context, approaches aimed at discovering spatio-temporal patterns are particularly relevant. This paper[1] focuses on spatio-temporal data and associated data mining methods.

1 Spatio-temporal Data

In recent years, technological advances in data acquisition (satellite images, sensors, etc.) have enabled numerous applications in surveillance and environmental monitoring: detection of abrupt changes (natural disasters, etc.), evolution tracking of natural phenomena (coastal erosion, desertification, wildfires, etc.) or development of models (hydrology, agriculture, etc.). The collected data is usually heterogeneous, multiscale, spatial and temporal (time series of satellite images, aerial or terrestrial photos, digital terrain models, physical ground measurements, qualitative observations, etc.). This data is used to understand and predict phenomena generated by processes that are complex and of multidisciplinary origin (climatic, geological, etc.). Exploitation by experts of those huge volume of complex data (big data) requires not only to structure it to the best but also and mainly to design data analysis and knowledge discovery methods. In that context, approaches involving pattern mining are particularly relevant.

With the dramatic growth of spatial information and Geographic Information Systems (GIS), many studies have been carried out in the context of spatiotemporal patterns mining. Early work in this area has dealt with spatial and temporal dimensions separately. Extraction of temporal sequences aims at identifying features frequent over time without taking into account spatial relationships. Colocation mining methods extract set of features which frequently appear in close objects without taking into account the temporal aspect. More recently, these works have been extended to simultaneously integrate spatial and temporal dimensions. Examples include the detection of sequences of located events and trajectory mining. A review has been published by the consortium GeoPKDD (Giannotti and Pedreschi, 2008). However, in those approaches, the mined patterns do not match the spatial complexity encountered when dealing with sattelite images. Similarly, the primitive constraints usually used (typically minimum frequency) are not sufficient to express criteria of interest for experts, such as geologists.

A spatiotemporal database contains information characterized by a spatial and a temporal dimensions. Two types of spatiotemporal databases are mainly considered: databases containing trajectories of moving objects located in both space and time (e.g. bird or aircraft trajectories); databases storing spatial and temporal dynamics of events (e.g. erosion evolution in a region or epidemic spread in a city).

2 Mining moving object trajectories

The emergence of new mobile technologies has facilitated the collection of large amounts of spatiotemporal data, dedicated to the localization of mobile objects in space and time (Perera et al., 2015). These new databases provide opportunities for new applications. The project GeoPKDD (Giannotti and Pedreschi, 2008), for example, studied

[1]The content of the paper was prepared in collaboration with H. Alatrista Salas, S. Bringay, F. Flouvat, and N. Selmaoui.

the development of traffic planning in large cities according to vehicle- flows. Other application domains include socio-economic geography, sports (e.g. football players), fishing control and weather forecast- (e.g. hurricanes). In most of these applications, the number of paths is high. One of the objectives of trajectory analysis is to find the most relevant paths according to the targeted problem (e.g. the most frequent, the most unexpected, periodic, etc.). Several approaches have been recently proposed in the literature, for instance (Orakzai et al., 2015).

3 Spatial patterns and spatiotemporal patterns for located event- mining

The extraction of spatial and spatiotemporal patterns has been studied extensively in recent years in geographic data and GIS. There are two families of approaches: colocations (Shekhar and Huang, 2001) that identify events that are frequently close; and spatiotemporal patterns that identify the evolution of events in both space and time (Alatrista-Salas et al., 2016). Sequences and more generally graphs have often been used and extended to the spatiotemporal context in order to represent the propagation of phenomena in space and time. Collocations focus on objects and their spatial relationships, for instance (Shekhar and Huang, 2001; Celik et al., 2008).

4 Conclusion

The challenges associated with spatial and spatio-temporal databases are numerous. Firstly, the semantics of extracted patterns must be considered to present experts with patterns which actually meet their application needs. Patterns with more complex structures, such as attributed graphs, can be really effective in spatial databases as shown by Pasquier's promising work (Pasquier et al., 1998) and (Sanhes et al., 2013). In addition, methods of spatio-temporal data mining often generate a lot of patterns, sometimes more than the size of original data. It is therefore important to define measures of interest that enable experts to select the most relevant patterns. As highlighted in the method based on colocations, it is also necessary to include - the domain knowledge (e.g. metadata, semantic descriptions, ontologies, etc.) in the extraction process to improve the scalability as well as the quality of the extracted patterns and their interpretation. A definition of relevant visualizations for those patterns would further facilitate their interpretation. Many application areas remain to be explored as for example image-mining where large amounts of data are available but few effective and scalable methods have been developed so far. Finally, there is a real need for collaboration between domain experts and data mining experts. Collaboration is the key to success for the knowledge extraction process.

References

Hugo Alatrista-Salas, Sandra Bringay, Frédéric Flouvat, Nazha Selmaoui-Folcher, and Maguelonne Teisseire. 2016. Spatio-sequential patterns mining: Beyond the boundaries. *Intell. Data Anal.*, 20(2):293–316.

Mete Celik, Shashi Shekhar, James P. Rogers, and James A. Shine. 2008. Mixed-drove spatiotemporal co-occurrence pattern mining. *IEEE TKDE*, 20(10):1322–1335.

Fosca Giannotti and Dino Pedreschi, editors. 2008. *Mobility, Data Mining and Privacy - Geographic Knowledge Discovery*. Springer.

Christian S. Jensen, Markus Schneider, Bernhard Seeger, and Vassilis J. Tsotras, editors. 2001. *Advances in Spatial and Temporal Databases, 7th International Symposium, SSTD 2001, Redondo Beach, CA, USA, July 12-15, 2001, Proceedings*, volume 2121 of *Lecture Notes in Computer Science*. Springer.

Faisal Orakzai, Thomas Devogele, and Toon Calders. 2015. Towards distributed convoy pattern mining. In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems, Bellevue, WA, USA, November 3-6, 2015*, pages 50:1–50:4.

Nicolas Pasquier, Yves Bastide, Rafik Taouil, and Lotfi Lakhal. 1998. Pruning closed itemset lattices for associations rules. In *BDA'98*.

Kushani Perera, Tanusri Bhattacharya, Lars Kulik, and James Bailey. 2015. Trajectory inference for mobile devices using connected cell towers. In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems, Bellevue, WA, USA, November 3-6, 2015*, pages 23:1–23:10.

Jérémy Sanhes, Frédéric Flouvat, Claude Pasquier, Nazha Selmaoui-Folcher, and Jean-François Boulicaut. 2013. Weighted path as a condensed pattern in a single attributed DAG. In *IJCAI 2013, Beijing, China, August 3-9, 2013*, pages 1642–1648.

Shashi Shekhar and Yan Huang. 2001. Discovering spatial co-location patterns: A summary of results. In Jensen et al. (Jensen et al., 2001), pages 236–256.