Compression-based Pattern Mining from Spatio-Temporal Graphs

Spatio-temporal data is generated ubiquitously as a result of many of our daily activities. For example, movement of humans in urban settings, cars over road networks, or check-ins in location-based social networks or public transportation systems. Spatio-temporal graphs can be used to model such data in terms of interactions of moving entities and spaces over time.

How can we automatically summarize and represent mobility patterns from spatio-temporal graphs? The minimum description length (MDL) principle states that the best summary can be found by finding the model that best compresses the data. The challenge of using MDL for spatio-temporal data is finding the proper encoding of spatial and temporal attributes. Examples of temporal parameters are delay, duration, and periodicity; example spatial parameters are distance topology, geometry, and spatial gravity.

The goal of this research is to come up with encodings of spatio-temporal connectivity behaviors using the MDL principle. Furthermore, we would like to use that encoding both for pattern mining and ranking the discovered patterns of the spatio-temporal graph in an efficient way. The developed algorithms will be tested on various spatio-temporal graphs, such as taxi trajectories and location-based social networks.

Example research questions that could be answered are:

- What spatio-temporal attributes are important and how can we encode the spatio-temporal attributes using the MDL principle?
- How can we make such a model robust to factors that degrade the quality of spatio-temporal datasets, (e.g. noise, missing data, etc.)?
- How can we rank and represent patterns found using the MDL principle?

Student profile
Good understanding of data mining and algorithms; programming in Python

Supervisors
Mitra Baratchi, Matthijs van Leeuwen

Relevant literature
Compression temporal graphs using MDL

[1] Shah, Neil, et al. "TimeCrunch: Interpretable dynamic graph summarization." *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2015.

Modeling spatial trajectories using MDL

[2] Lee, Jae-Gil, Jiawei Han, and Kyu-Young Whang. "Trajectory clustering: a partition-and-group framework." *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*. ACM, 2007.

Methods on spatio-temporal data mining

[3] Zheng, Yu. "Trajectory data mining: an overview." *ACM Transactions on Intelligent Systems and Technology (TIST)* 6.3 (2015): 29.