Trajectory Data Mining in the Age of Big Data and AI

Manos Papagelis

Electrical Engineering and Computer Science
Lassonde School of Engineering
York University
papaggel@eecs.yorku.ca

1 Introduction

Trajectory Data. *Trajectory data* consists of records that capture the movement of objects or entities over time, typically provided in the form of triplets $\langle o, t, (x, y)\rangle$, representing that an object o at timestamp t was at location with coordinates (x, y). The proliferation of location-based technologies, geo-enabled smart devices, and advanced global positioning systems has resulted in the accumulation of vast amounts of trajectory data.

Trajectory Data Mining. *Trajectory data mining* focuses on extracting valuable patterns, information, and insights from trajectory data, and it has been an active research direction for a long time [1, 2, 3]. Mining interesting patterns and extracting useful information from trajectories can find application in diverse domains, including intelligent transportation systems, urban planning and environmental monitoring, and public health. Due to the broader impact, several trajectory-related research problems have been of interest, including trajectory similarity, prediction, clustering, classification, simplification, outlier detection, and imputation.

2 Our Research Contributions

Over the last years, we have been revisiting classic trajectory data mining problems through the lens of modern technologies and methods, including *big data analysis*, *graph mining / network analysis*, and *deep learning methods*. Our journey includes methods for:

- **Trajectory simplification** (ACM SIGSPATIAL '23): We presented PathletRL, a deep reinforcement learning method for identifying a small set of trajectory building blocks, known as *pathlets*, that can compactly represent a vast number of trajectories.
- **Generating higher-order trajectory data** (ACM SIGSPATIAL '23). We presented Point2Hex, a method and tool for generating higher-order mobility flow datasets from raw trajectory data.
- **Trajectory-user linking** (IEEE MDM '23): We presented tulhor, a Transformer-based model that links anonymous trajectories to the respective users.
- **Trajectory network analysis** (IEEE MDM '18; GeoInformatica, 23, '19; IEEE BigData '18; IEEE MDM '20): We presented methods for modeling interactions of moving objects using graphs, including methods for (*i*) trajectory node centrality computation and for (*ii*) mining pedestrian group dynamics.
- **Transportation Optimization** (ACM SIGSPATIAL '22, ACM SIGSPATIAL '22): We presented methods for (*i*) forecasting the performance of road intersections, and for (*ii*) the vehicle navigation problem.
- **Mobility and Epidemics** (ACM SIGSPATIAL/SpatialEpi '23, ACM SIGSPATIAL/SpatialEpi '23, IEEE MDM '22): We presented methods for modeling epidemic spreading in mobility networks.

3 Presentation Structure

In this presentation we will delve into our recent endeavors on trajectory data mining, emphasizing its contemporary and evolving nature that stems from the incorporation of novel deep learning methods for addressing long-established problems. Depending on time, we will present our recent work on trajectory-user linking, a trajectory classification problem aimed at connecting anonymous trajectories to their respective users (IEEE MDM '23). We will also present our recent work on trajectory dictionary construction, which aims at constructing a trajectory pathlet dictionary (ACM SIGSPATIAL '23).

References

[1] Gowtham Atluri, Anuj Karpatne, and Vipin Kumar. Spatio-temporal data mining: A survey of problems and methods. *ACM Comp. Surveys*, 51(4), Aug 2018.

[2] Ali Hamdi et al. Spatiotemporal data mining: a survey on challenges and open problems. *Artificial Intelligence Review*, 55(2):1441–1488, Feb 2022.

[3] Yu Zheng. Trajectory data mining: An overview. *ACM Trans. Intell. Syst. Technol.*, 6(3), May 2015.