Generating Synthetic Spatio-temporal Car-hailing Traffic Knowledge Graph

Supervisors:
- Jiancheng Weng (Professor, BJUT, youthweng@bjut.edu.cn)
- Ronald Siebes (Assistant Professor, VU Amsterdam, r.m.siebes@vu.nl)
- Shuai Wang (Scientific Engineer, VU Amsterdam, shuai.wang@vu.nl)
- Zhisheng Huang (Emeritus Professor, VU Amsterdam, huang.zhisheng.nl@gmail.com)

Student: Mateusz Grzegorz Kędzia, a student in the joint Master's programme in Artificial Intelligence (a joint degree by VU and UvA, m.g.kedzia@student.vu.nl).

Exchange Period: 1st March to 31st August 2025

Thesis Proposal

As one of the most populated cities in the world, Beijing has one of the most active car-hailing services with approximately 55 thousand drivers driving 60 thousand cars from 12 companies as of the end of 2023[1]. This accumulates a significant amount of data that could be used for the analysis of vehicle trajectories. Huang et al. proposed to use LSTM-AE-Attention for the detection of a large amount of trajectories of GPS [1]. Their model has an F1 score is 9.8% higher than the average F1 score of the other selected models in comparison for the task of the abnormal GPS trajectory detection. Another related research has been conducted by He et al. using taxi data in Shanghai and Beijing with a 10% increase in the classification of trajectories [2]. Some other interesting works include an extension of the multi-scale convolutional neural network model (MCNN) with the time series symbolization algorithm (TSA). This hybrid approach achieved better performance in behavior classification including normal driving, speeding driving, emergency stopping, and temporary stopping than the models in comparison [3]. Existing benchmarks and extracted features are either outdated or not at the scale of the entire city. Moreover, this traffic data of car-hailing remains not accessible to the public due to various reasons. In this project, we aim to study the generation of synthetic data that matches the real data on selected indicators, constraints, and graph-theoretical metrics. As a use case, we apply existing abnormal vehicle trajectory detection algorithms and benchmark the performance on real car-hailing traffic data in Beijing and our synthetic data. If possible, we would like to also include some comparison with taxi data.

The student will generate, analyze, and compare the graphs in the following four stages:

1. First, the student wil construct a spatio-temporal knowledge graph by integrating provided traffic data sources [4, 5]. After filtering out the abnormal trajectory data using the methods [1, 2], the student will enrich it with locational information by

[1] Retrived from the official website of the Beijing Municipal Commision of Transport: https://jtw.beijing.gov.cn/czqc/hygk/202110/t20211021_2517443.html#:~:text=%E7%BD%91%E7%BA%A6%E8%BD%A6%E5%B9%B3%E5%8F%B0%E4%BC%81%E4%B8%9A,%E5%90%8C%E6%AF%94%E4%B8%8A%E5%B9%B4%E5%A2%9E%E5%8A%A019.9%25%E3%80%82.

associating the geolocation with corresponding entity type information such as hotels, airports, schools, etc. The corresponding information could be found/inferred from open source data from online repositories.

2. The student will extend and customize existing spatio-temporal knowledge graph generation algorithms with means of maintaining the above-mentioned type of information and perform refinement on the generated graphs. Candidate algorithms to consider include the GAN-based approach [1], and other methods [7] depending on the data provided. The generated graphs will be evaluated against selected traffic indicators (e.g. temporal distribution, trajectory density, trajectory distance) and constraints (density constraints, distance constraints) [8], as well as well-known graph-theoretical metrics (e.g. centrality, closeness).
3. As a use case, the student will perform trajectory classification on both data by applying two methods: a) the LSTM-AE-Attention deep learning model for vehicle trajectory reconstruction and anomalous trajectory detection proposed by Huang et al. [1] and b) a method based on the Specific Support Vector Machine (SSVM) by He et al. [2] with the possibility to extend to other methods [3].
4. The project concludes with some comparative analysis with reflection on the limit of the graph generation algorithms and the abnormal behavior detection of the algorithms on real data, including speeding driving, emergency stopping/starting, temporary stopping, taxi frauds [3, 6]. Possible extensions of this project includes adapting the method and feature extract to taxi data.

The novelty and research value of this project lies in the following aspects

- We extend existing algorithms mentioned above and explore possible means of generating large geo-spatial knowledge graphs while preserving the semantics: the nodes are associated with types (e.g. hotel, airport, tram stations) and such types could be inferred from their geolocation information or external sources of open data.
- We compute generic traffic indicators for both the pre-processed data as well as synthetic data for the evaluation of the graph-generating algorithm.
- We perform abnormal trajectory detection on both graphs, report the similarity, and discuss possible future steps to improve the algorithm to achieve better similarity.
- The resulting synthetic data serves as a unique large-scale data source for comparative studies with other megacities.

Data privacy and sensitive data protection

- The original data is to be provided by Prof. Jiancheng Weng. The original data will be anonymized with sensitive information removed before the student gets involved in the project. Additional data resources used for enrichment are going to be open data.
- The pre-processed car-hailing traffic data, the synthetic data, as well as all the data generated in intermediate steps will only be stored and processed on servers in the BJUT and will not leave BJUT during the project. The supervisors from VU Amsterdam will not have access to the pre-processed car-hailing traffic data.
- We would consider publishing only the synthetic data and essential intermediate results after the project with permission from Prof. Weng and related stakeholders.
- The student will be granted access to the data from 6th March 2025. The last day for the student to work on this project is 31st August 2025. The student will no longer have access to the data after the internship.

The project will start on 6th March 2025. The student is expected to have the thesis draft ready by 1st July. The defense is expected no later than 1st August 2025. The final version of the thesis should be submitted no later than 14th August 2025. The mark of this project is going to be submitted the the admission system of the Vrije Universiteit Amsterdam as the final grade of the student's master's thesis project, which is used to validate his master's degree requirement.

References:

[1] Shi-chen Huang, et al., Vehicle Trajectory Reconstruction and Anomaly Detection Using Deep Learning, Journal of Transportation Systems Engineering and Information Technology, 2021

[2] Ming He, et al., Abnormal trajectory detection method based on enhanced density clustering and abnormal information mining, Journal on Communications, 2017

[3] Jian-dong Zhao, et al., Recognition of Abnormal Driving Behavior of Key Commercial Vehicles, Journal of Transportation Systems Engineering and Information Technology, DOI:10.16097/j.cnki.1009-6744.2022.01.030.

[4] Guo Xuan, et al., A Method for Constructing Geographical Knowledge Graph from Multisource Data. https://www.mdpi.com/2071-1050/13/19/10602

[5] JIANG Bingchuan, et al. Geographic Knowledge Graph Building Extracted from Multi-sourced Heterogeneous Data, Acta Geodaetica et Cartographica Sinica. http://xb.chinasmp.com/EN/Y2018/V47/I8/1051

[6] HAN Boyang, et al., An anomaly detection algorithm for taxis based on trajectory data mining and online real-time monitoring, Journal of University of Science and Technology of China, 2016.

[7] Kong, X., Chen, Q., Hou, M. et al. Mobility trajectory generation: a survey. Artif Intell Rev 56 (Suppl 3), 3057–3098 (2023). https://doi.org/10.1007/s10462-023-10598-x

[8] Quan Liang, Research on Travel Behavior Forcasting Method of Public Transport Commuters Based on Individual Travel Graphs, PhD thesis, Beijing University of Technology, 2019.