JOHANNES KEPLER
UNIVERSITY LINZ

Univ.-Prof.in Dr.in
Cristina Olaverri Monreal
Lehrstuhlinhaberin Nachhaltige
Transportlogistik 4.0

T +43 732 2468 5490
cristina.olaverri-monreal@jku.at
www.jku.at/its

Sekretariat:
Oliver Bartenstein
DW 5491
oliver_martin.bartenstein@jku.at

Master Thesis:
Generative Modeling for Transportation Data

Table of Contents

I. Research Idea

Various innovative and disruptive solutions in the field of intelligent transportation systems (ITS), such as autonomous driving and mobility analytics, involve making predictions based on data. Furthermore, generating such predictions is highly non-trivial given the typical sensitivity of the system to small perturbations such as traffic incidents, weather conditions, special events, etc.

At the intersection of machine learning and statistics, **generative modeling** can be used as a promising method to analyse, predict, and generate spatio-temporal mobility patterns. Generative modeling focuses on estimating the joint probability distribution of data, which can be used in turn for a variety of tasks such as generating synthetic data, assessing quality of simulated data, imputing data samples where necessary (e.g. missing data), and quantifying uncertainty of predictions.

In recent years, generative models such as generative adversarial networks (GANs) allowed considerable progress in large scale problems such as computer vision (image data) and natural language processing (text data). The goal of this thesis will be to extend the application of generative modeling for other types of high-dimensional data such as tabular data and time series in the field of transportation modeling.

II. Scientific relevance of the research idea compared to related published work

Research and development in the field of ITS highly depends on good quality data and robust predictions, not only for ensuring optimized solutions, but also for safety reasons. In many scenarios, there is a lack of real-world data and/or missing/noisy data, hence the interest of generating “realistic” data using generative models. These models can also be used to analyze simulated data and generate predictions for mobility patterns. Generative models have been extensively applied to high dimensional data such as images (e.g. generating images of realistic faces based on input images of real faces) or text (e.g. completing sentences, generating a body of text that seems realistic). It is however challenging to use them based on data of a smaller scale such as tabular transportation data (floating car data, sensor data) from a limited time span. This thesis can explore this direction.

III. Goal of the thesis

- Study the existing literature related to generative modeling for transportation
- Apply generative modeling to predict mobility patterns based on various available transportation datasets. The considered datasets can be either openly available or datasets that are already available at the Chair.
- Apply generative modeling to assess quality of simulated data
- Develop an uncertainty-aware generative model to create synthetic data that is useful in a transportation problem
- Possibly contribute to a novel research publication based on the work done for the thesis

IV. Own research questions and hypotheses

- Under which conditions are generative models applicable and useful for transportation datasets? e.g. how much data is sufficient to achieve good results?
- How to assess quality of generated data, e.g. generated trajectories compared to real trajectories?
- How to quantify uncertainty of predictions for a transportation scenario?
- How to integrate different input sources, e.g. maps, gps data, landmarks?

V. Short presentation of the planned method

The planned method can be a generative adversarial network (GAN). It will however depend on the considered scenario, input data, and the student's initiative. The student will be highly encouraged to develop a novel method.

VI. Thesis structure design

The thesis will have a typical structure consisting in the following sections: Introduction, related work, methodology, results and discussion, and conclusion.

VII. Draft of time plan

The presented thesis is planned for a six-month period. An initial time plan:

- **1st month**: Study related literature and identify opportunities to apply generative modeling for transportation problems.
- **2nd-4th month**: Apply generative modeling to transportation datasets, present and refine results according to discussions with the supervisor. Identify a problem of interest and develop a novel method based on generative modeling.
- **5th-6th month:** Write master thesis and possibly contribute to an original research paper.