DTU MLSM
Machine Learning for Smart Mobility

MSc Thesis Themes

This document contains research topics for potential MSc thesis with the MLSM group (http://mlsm.man.dtu.dk). They are organized by 12 different themes (see list below). Within each theme, you will find different research topics (bullets) that we are interested in exploring in the context of a MSc thesis.

If you are interested in any of these topics, please fill out the following form: https://forms.gle/9yH6yiKumfP229S38

Index of MSc thesis themes (click to read more):

1. Machine Learning for Mobility-on-Demand

BACKGROUND: Mobility-on-Demand (MoD) is an innovative concept in the transportation field where the mobility service is provided on a request from a customer. Nowadays, we witness a growing number of companies operating in Denmark and worldwide which are based on this concept, such as SHARENOW, Green Mobility or Donkey Republic bikes. The shared mobility services provided by these companies allow for efficient utilization of transport networks and vehicles hence contributing to sustainable development of modern cities. However, being highly responsive to customer needs, this business model more than ever relies on accurate prediction of short- and long-term travel demand, defining business areas for further expansion, fleet rebalancing, etc.

POTENTIAL RESEARCH TOPICS:

- Modelling the dynamics of the spatial distribution of demand through time (see e.g. the application in https://arxiv.org/abs/2006.05256)
- Predicting time-to-pickup of cars/bikes given drop-off time and characteristics of the drop-off location and other contextual information (number of nearby cars/bikes and their charging/fuel status, distance to public transport, points-of-interest, etc.)
- Deep spatio-temporal models of urban mobility - since transportation demand is correlated across time and space, the goal is to build models that can exploit those correlations to improve predictions (see e.g. https://link.springer.com/article/10.1007/s00521-018-3470-9)
- Zero-shot learning of shared mobility demand - the goal is to explore meta learning methods (e.g. https://arxiv.org/abs/1703.03400) to quickly adapt demand prediction models to new areas and cities
- Censored models for estimating latent true demand for cases when supply is limited (see e.g. https://arxiv.org/abs/2001.07402 and https://arxiv.org/abs/2009.04822)
- Reinforcement learning for rebalancing Mobility-on-Demand supply

POTENTIALLY RELEVANT METHODOLOGIES: Graph Neural Networks, Stochastic Recurrent Neural Networks, Survival Analysis, Normalizing Flows, Reinforcement Learning, Meta Learning, Zero-shot Learning

CONTACTS:

- Associate Prof. Filipe Rodrigues [rodr@dtu.dk]
- Prof. Francisco Pereira [camara@dtu.dk]
- Associate Prof. Carlos Azevedo [climaz@dtu.dk]

2. Understanding how people make choices through Machine Learning

BACKGROUND: Understanding and predicting the behavior of humans has long been one of the holy grails of behavioral science, with tremendous impact in many areas such as policy making for sustainability, energy systems, transportation, healthcare, urban design, and marketing. Consider, for example, the introduction of a new, more sustainable, transportation alternative. It is crucial to understand how people will change their mobility choices in response to the new alternative in order to properly adjust its characteristics (pricing, frequency, capacity, zoning, etc.) and assess its overall feasibility. However, predicting behavior is notoriously challenging. Humans have varying preferences, motivations, experiences and decision processes. Therefore, in order to understand and predict their complex behavior, we require a flexible statistical framework that is able to capture rich behavioral processes and associated uncertainties, while having inference procedures that scale to very large datasets.

POTENTIAL RESEARCH TOPICS:

- Scaling Bayesian inference in choice models to very large datasets and improving out-of-sample generalization using deep neural networks and amortized variational inference (see e.g. https://arxiv.org/abs/2004.05426)
- Automatic utility function specification in large-scale choice models such as mixed-logit models (see e.g. https://arxiv.org/abs/1906.03855)
- Bayesian online learning for choice models to allow them to work with streaming data and in a distributed manner
- Flexible and interpretable choices models by combining discrete choice models with machine learning methods (see e.g. https://arxiv.org/abs/2101.12252 and https://arxiv.org/abs/2007.02739)

POTENTIALLY RELEVANT METHODOLOGIES: Bayesian Modelling, Deep Learning, Discrete Choice Models, Representation Learning, Bayesian Nonparametrics, Variational Inference

CONTACTS:

- Associate Prof. Filipe Rodrigues [rodr@dtu.dk]
- Prof. Francisco Pereira [camara@dtu.dk]

3. Reinforcement Learning for Transportation

BACKGROUND: Reinforcement Learning (RL), in particular it's combination with Deep Learning, is currently one of the most exciting and promising research areas in AI. RL has recently shown outstanding results at solving complex sequential decision-making problems such as learning to play Go and videogames at super-human-level performance, autonomous driving and smart grid optimization. Through interaction with an environment (real or simulated), RL algorithms are able to learn optimal policies for taking actions at each time step t, such that the sum of expected future rewards is maximized (e.g. maximizing game score in the videogame example). The ability of RL algorithms to correlate actions with the delayed rewards that they lead to (e.g. amount of congestion reduction, or responsiveness increase) makes them excellent candidates for learning policies for complex transportation challenges, such as adaptive autonomous vehicle routing and re-balancing through dynamic pricing, traffic signal control for reducing congestion, etc.

POTENTIAL RESEARCH TOPICS:

- Reinforcement learning for rebalancing Mobility-on-Demand supply
- Meta reinforcement learning for traffic signal control (see e.g. https://arxiv.org/abs/1904.08353 and https://dl.acm.org/doi/10.1145/3447556.3447565)
- Model-based reinforcement learning for dynamic pricing of transportation services and network usage (see e.g. https://www.ijcai.org/Proceedings/2019/0635.pdf)
- Cross-area and cross-city transferability of learned RL policies (e.g. for traffic signal control, rebalancing of shared mobility services or autonomous vehicle routing)
- New reinforcement learning and meta reinforcement learning methodologies

POTENTIALLY RELEVANT METHODOLOGIES: Deep Reinforcement Learning, Graph Neural Networks, Bayesian Deep Learning, Meta Learning, Deep Generative Models, Model Predictive Control

CONTACTS:

- Associate Prof. Filipe Rodrigues [rodr@dtu.dk]
- Associate Prof. Carlos Azevedo [climaz@dtu.dk]

4. Combining Simulation and AI

BACKGROUND: Simulation models are usually very complex and detailed tools that can represent cause-effect interactions in space and time based on rules of the domain (e.g. physics, human behavior, biology, climate). For that reason, they are still today an unavoidable tool for many large scale problems, such as transport and urban modeling, climate change, drug development. On the other hand, they are known to be extremely challenging tools to work with: they are often very slow to run, difficult to calibrate, require a tremendous amount of memory, and are hard to reuse.

The goal of this theme is to apply Machine Learning methods towards the next generation of Simulation research, with automatic calibration, efficient exploration of the parameter space, and better generalizability.

Significant advancements in this theme will have serious impact in real-world policy problems, such as climate adaptation and mitigation, green transition, energy policy.

POTENTIAL RESEARCH TOPICS:

- Automatic calibration - learn the best parameters for the simulator given real data observations (see. e.g. https://www.sciencedirect.com/science/article/pii/S1877050918304605)
- Scenario discovery - exploring the gigantic space of parameters from a simulator, to efficiently select the most interesting/promising scenarios (see e.g. https://ideas.repec.org/a/eee/tefoso/v111y2016icp124-134.html)
- Metamodeling - learning an efficient ML function that approximates the inefficient simulation model (see e.g. https://www.sciencedirect.com/science/article/pii/S1569190X19300802)
- Agent based modelling and simulation- large-scale simulation of complex systems (mobility, energy, etc) using combined ABM, network and ML techniques (see e.g. https://ieeexplore.ieee.org/document/8569541)

POTENTIALLY RELEVANT METHODOLOGIES: Active Learning, Bayesian Optimization, Deep Reinforcement Learning, Graph Neural Networks, Dynamic Traffic Assignment, Agent-based modelling, Activity Based Modeling, Complex Networks

CONTACTS:

- Associate Prof. Carlos Azevedo [climaz@dtu.dk]
- Prof. Francisco Pereira [camara@dtu.dk]
- Associate Prof. Filipe Rodrigues [rodr@dtu.dk]

5. Causality in Machine Learning for Sustainable Modeling and Prediction

BACKGROUND: Causality is arguably one of the major challenges of AI today, because the largest part of existing statistical methods relies on finding correlation patterns in the data. In other words, these models are correlational but ***not*** causal, which makes them less capable in making predictions out of their original training distribution (a.k.a. zero-shot problems; out-of-distribution test sets; external validity scenarios). A very important example of this is predicting the impacts of policies today in future climate and sustainability scenarios.

Recent research in Machine Learning (e.g. from Judea Pearl, Elias Bareinboin, Jonas Peters, Bernard Schölkopf, David Blei and many others) proposes different methods to apply causality in data-driven decision making. Quite often, these methods rely either on strong knowledge of the domain (we should know the *causal graph*), or on access to complex experimental data (e.g. randomized controlled trials), and the latest trends (e.g. Graph Neural Networks, Reinforcement Learning, Probabilistic Graphical Models, Physics Informed Deep Learning) promise to push for exciting possibilities using general observational data.

POTENTIAL RESEARCH TOPICS:

- Causal discovery from simulation models - Automatic extraction of the underlying causal graph(s) by repeatedly inspecting a simulation model
- Causal reinforcement learning (e.g. https://stjohngrimbly.com/causal-reinforcement-learning/)
- Causal discovery and causal representation learning from observational data (e.g. https://ieeexplore.ieee.org/abstract/document/9363924)
- Causal inference (e.g. https://arxiv.org/abs/2008.07283)

POTENTIALLY RELEVANT METHODOLOGIES: Graph Neural Networks, Reinforcement Learning, Probabilistic Graphical Models, Physics Informed Deep Learning, Active Learning

CONTACTS:

- Prof. Francisco Pereira [camara@dtu.dk]

6. Control of Multimodal Mobility Systems

BACKGROUND: The nature of the mobility system today includes a wide range of decision makers and technology, giving rise to different business models, services and solutions. Cities now often offer bus, metro, rail, car-sharing, on-demand, bike sharing or personal mobility solutions in a large mesh of technological and operating settings. These services tend to operate autonomously, while new opportunities for information sharing and coordinated actions emerge with new technologies. This research line focus on new methodologies for achieving a more efficient mobility system as a whole, both from theory and solution perspectives.

POTENTIAL RESEARCH TOPICS:

- Bayesian Optimization for Congestion Pricing (see e.g. https://arxiv.org/abs/2012.11047)
- Multi-agent decision making for autonomous mobility in multimodal systems (see https://www.sciencedirect.com/science/article/abs/pii/S0965856420306133)
- Tradable permits in mobility (see https://arxiv.org/abs/2101.00669) and emissions (see https://ieeexplore.ieee.org/abstract/document/9264892)

POTENTIALLY RELEVANT METHODOLOGIES: Bayesian Optimization, Reinforcement Learning, Simulation-based Optimization, Multi-Agent Systems, Control Theory and Economics (pricing).

CONTACTS:

- Associate Prof. Carlos Azevedo [climaz@dtu.dk]
- Prof. Francisco Pereira [camara@dtu.dk]
- Associate Prof. Filipe Rodrigues [rodr@dtu.dk]

7. AI and Simulation for Impact Assessment

BACKGROUND: While AI and advanced simulation techniques are emerging in the mobility planning and operations domain, they are also extensively being explored in related domains such as health, energy, emissions and climate impacts of mobility and transport.

POTENTIAL RESEARCH TOPICS:

- 2-way assessment of mobility and health impacts relationship, e.g.:
  - Urban environment vs. mental health
  - Mobility patterns vs. physical health
- Energy
  -
- Emissions
  - Mobility patterns vs. pollution exposure (especially for cyclists)
- Climate

POTENTIALLY RELEVANT METHODOLOGIES: Simulation, Active Learning, Bayesian Optimization, Deep Learning, Probabilistic Graphical Models.

CONTACTS:

- Associate Prof. Carlos Azevedo [climaz@dtu.dk]
- Prof. Francisco Pereira [camara@dtu.dk]

8. Generative Models and Uncertainty

BACKGROUND: In the real world, data is full of uncertainty, coming from inherent noise, limitations in measurement technology, and random external disturbances (e.g. disruption events). For this reason, in order to make useful predictions for decision making, one often needs to model entire distributions (as opposed to the *mean*). This allows for understanding how "certain" is a prediction, what would a "best case" and a "worst case" scenario look like, are there different particular values with high probability?

All of this is made possible through generative modeling tools that come into ML through the Bayesian framework. In this theme, we explore this framework to address different kinds of problems that involve uncertainty, complex spatial-temporal relationships, and high dimensionality.

POTENTIAL RESEARCH TOPICS:

- Predicting spatio-temporal distribution of transportation demand and its dynamics through space and time (see e.g. the application in https://arxiv.org/abs/2006.05256)
- Modelling uncertainty in bus arrival time prediction by exploring Bayesian deep learning and deep quantile regression (see e.g. https://arxiv.org/abs/1808.08798 but applied to https://arxiv.org/abs/1903.02791)
- Spatial conditional density estimation with normalizing flows for modelling traffic incident data, bike accident data or crime data (see e.g. https://arxiv.org/abs/1802.04908 and https://arxiv.org/abs/1903.00954)
- Deep generative models and active learning methods for electric vehicle (EV) charging network design and optimization
- Learning to simulate human mobility (see e.g. https://dl.acm.org/doi/10.1145/3394486.3412862)
- Population synthesis - how to generate a new "virtual" dataset that reflects the same distributional characteristics as the original one, but preserves privacy and respects domain constraints (see e.g. https://arxiv.org/abs/1808.06910)

POTENTIALLY RELEVANT METHODOLOGIES: Graph Neural Networks, Physics-informed Deep Learning, Heteroscedastic Models, Active Learning, Variational Autoencoders, Normalizing Flows, Generative Adversarial Networks

CONTACTS:

- Prof. Francisco Pereira [camara@dtu.dk]
- Associate Prof. Filipe Rodrigues [rodr@dtu.dk]

9. Deep Spatio-Temporal Models of Transportation Supply

BACKGROUND: In recent years, road congestion has been increasing in Denmark and around the world, and a rapid increase is also expected in the coming years. Road congestion is a considerable nuisance to travelers and, importantly, constitutes a large annual societal loss. Furthermore, the additional environmental impact is considerable, and in general more than 25% of the Danish emissions stem from road transport, and this share is increasing. In order to be able to limit the consequences of congestion and reduce the environmental impact, it is important that the current road network is utilised as well as possible and that long term planning of new infrastructure investments are supported by realistic prognosis of their impacts. This research theme aims at contributing to that.

POTENTIAL RESEARCH TOPICS:

- Graph neural networks for traffic speed/flow forecasting - since traffic congestion is correlated across time and space (the road network), the goal is to build models that can exploit those correlations to improve predictions (see e.g. https://ojs.aaai.org/index.php/AAAI/article/view/3877 and also https://ojs.aaai.org/index.php/AAAI/article/view/3881)
- Physics-informed deep learning for traffic prediction - the goal is to incorporate information about the physical properties of traffic dynamics (e.g. from traffic flow theory) into deep neural networks (see e.g. https://arxiv.org/abs/2101.06580)
- Bus arrival time prediction using graph neural networks to capture relations between multiple links in the bus network and propagating delays - based on previous work (https://arxiv.org/abs/1903.02791), the goal is to leverage recent advances in graph neural networks to improve predictions
- Few-shot learning of traffic congestion and travel time - the goal is to explore meta learning methods (e.g. https://arxiv.org/abs/1703.03400) to quickly adapt transportation prediction models to new areas and cities
- Predictive inference for travel time on transportation networks - goal is to understand how to aggregate travel-time prediction distributions over many segments to arrive at the distribution of travel time over a route (see e.g. https://arxiv.org/abs/2004.11292)

POTENTIALLY RELEVANT METHODOLOGIES: Stochastic Recurrent Neural Networks, Graph Neural Networks, Physics-informed Deep Learning, Traffic Flow Theory, Meta Learning, Few-shot Learning, Normalizing Flows

CONTACTS:

- Associate Prof. Filipe Rodrigues [rodr@dtu.dk]
- Prof. Francisco Pereira [camara@dtu.dk]

10. Next-generation Intelligent Transportation Systems

BACKGROUND: With this research stream we aim to explore advancements in technology in smart urban systems and the underlying theoretical methods for its operation. We focus on the interconnectedness of various smart urban systems, indicators and technologies. In particular we focus on the development of new algorithms and frameworks for the next generation of ITS, being it user-centric through smartphones or information coordination of self-driving vehicles.

POTENTIAL RESEARCH TOPICS:

- Mobility Tokens and Blockchain: Development of a Smart Mobility Management System (see https://arxiv.org/abs/1908.05629)
- Dynamic, Pricing, Incentives, and Tradable Permit Schemes (see https://arxiv.org/abs/2101.00669)
- Mobility-as-a-Service Systems (see https://discovery.ucl.ac.uk/id/eprint/10061860/7/Kamargianni_TRB2019_MaaSsim_Extended%20Abstract_full.pdf)

POTENTIALLY RELEVANT METHODOLOGIES: Block-chain, Simulation, Control Theory, Deep learning, Reinforcement Learning

CONTACTS:

- Associate Prof. Carlos Azevedo [climaz@dtu.dk]
- Prof. Francisco Pereira [camara@dtu.dk]
- Associate Prof. Filipe Rodrigues [rodr@dtu.dk]

11. Machine Learning Methodology

BACKGROUND: Besides the more application-oriented themes described above (and although many of them can become deeply theoretical), we are also interested in MSc thesis that are purely methodology-oriented and that don't directly consider an application in the transportation domain. Below is a list of machine learning fields that are of interest to us from a pure methodological research perspective. Theoretical advances in these fields are also extremely likely to result in important practical advances in transportation.

POTENTIAL RESEARCH TOPICS:

- Bayesian Deep Learning
- Amortized Variational Inference
- Normalizing Flows
- Reinforcement Learning
- Meta Learning
- Graph Neural Networks
- Zero-shot and Few-shot Learning
- Causal inference
- Causal representation learning
- Counterfactual reasoning
- Inference in Stochastic Recurrent Models
- Representation Learning
- Bayesian Optimization
- Symbolic AI

CONTACTS:

- Associate Prof. Filipe Rodrigues [rodr@dtu.dk]
- Prof. Francisco Pereira [camara@dtu.dk]

12. Visualization

BACKGROUND: As a tool, visualization can be much more than a pleasant way to communicate a statistical concept. It can be a method to gain insights, as much as many other ML methods can be, such as clustering or dimensionality reduction. In this theme, we invite students to work on new visualization methods for complex prediction problems.

A particular note in this case is that MLSM does not have itself strong visualization background, thus any thesis in this case will require either a co-supervisor that is expert in the area, or a student that is creative and technically strong in it.

POTENTIAL RESEARCH TOPICS:

- Visualization of predictions in space and time
- Visualization of uncertainty
- Visualization of model performance across different datasets
- ...