Australasian Transport Research Forum 2021 Proceedings
8-10 December, Brisbane, Australia
Publication website: http://www.atrf.info

Network-wide traffic simulation with multi-agent imitation learning

Jie Sun[1] and Jiwon Kim[1*]
[1] School of Civil Engineering, The University of Queensland, St. Lucia 4072, Brisbane, Australia
Email for correspondence: jiwon.kim@uq.edu.au

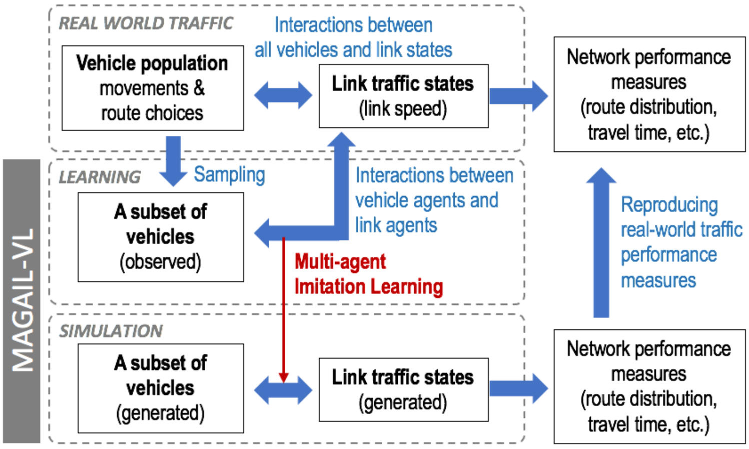
1. Introduction

Due to the high complexity of traffic system, traffic simulation is an essential and efficient approach for the analysis and evaluation of the traffic system by modelling traffic flow dynamics and vehicle mobility in response to information and control actions, where network-wide traffic simulation could enable traffic engineers to predict the spatio-temporal movement patterns of vehicles and develop network traffic management strategies to alleviate traffic congestion (Mahmassani, 2001, Kim and Mahmassani, 2015). However, building conventional traffic simulation models is often time-consuming due to complex parameter estimation and calibration processes needed for a high-fidelity simulation model. Therefore, data-driven simulation has gained considerable attention in the recent decade with the increasing availability of high-resolution vehicle trajectory data and massive advances in deep learning models. While numerous data-driven microscopic traffic models have been proposed in the last decade, there is little effort made on network-wide mesoscopic/macroscopic traffic simulation based on trajectory data (Li et al., 2020). In this paper, we aim to develop a model that leverages both high-resolution trajectory data and deep learning to learn interactions between vehicles and a road network, which can provide the basis for enabling data-driven mesoscopic traffic simulation at the network-level.

While it is a relatively new concept in the transport research community, several relevant research topics have been studied, including the next location prediction problem, which aims to predict the next location in a trajectory of a user based on the previously visited locations (Sun and Kim, 2021), and vehicle trajectory generation problem, which aims to generate synthetic trajectory data using generative models to learn the mobility patterns (Ziebart et al., 2008, Choi et al., 2021). However, these studies model the movements of individual vehicles independently without considering the interactions between vehicles and between vehicles and a traffic network, which leads to limited applicability.

Our study aims to model the interactions of vehicles and the road network by employing a *multi-agent imitation learning* (MAIL) framework, which learns behaviours of a multi-agent system based on demonstrations of a set of experts interacting with each other. By considering observed traffic and trajectory data 'expert demonstrations', the imitation learning (IL) approach can train a multi-agent model to generate trajectories and traffic effects that mimic the 'demonstrated' real-world behaviours. A naïve approach to building such a multi-agent model might be to model the whole vehicle population in the network as individual vehicle agents, but it is computationally expensive. Currently, the largest number of agents modelled in the literature of multi-agent imitation learning is only 100 (Bhattacharyya et al., 2018). As such, we propose the idea of learning the interactions between *a subset of vehicles* and *road link traffic states*, instead of attempting to model the interactions among all vehicles, by developing the **M**ulti-**A**gent **G**enerative **A**dversary **I**mitation **L**earning model with **V**ehicle and **L**ink agents (MAGAIL-VL). Figure 1 shows our model framework. In a real-world traffic network, vehicles' route choice actions determine the traffic states across the links (e.g., link speed) and the link traffic states, in turn, affect vehicles' route choices. The consequences of such interactions are captured through various network performance measures such as route distribution and travel times. Unlike traditional mesoscopic traffic simulation models, which typically use dynamic traffic assignment (DTA) to model route choice behaviour and vehicle-road link interactions, our MAGAIL-VL model attempts to learn the vehicle-road interaction patterns directly from data by applying MAIL to a subset of vehicle trajectory data to train vehicle agents and all link traffic data to train link agents. This learned interaction is then transferred to a simulation environment, where a new subset of vehicles and link states can be generated and simulated to produce the network performance measures that mimic the network performance under the whole population. The goal is to predict the population network performance measures by modelling only a subset of vehicles and their interactions with the underlying road network.

Figure 1: Schematic diagram of the concept of this study


REAL WORLD TRAFFIC
Interactions between all vehicles and link states
Vehicle population movements & route choices
Link traffic states (link speed)
Network performance measures (route distribution, travel time, etc.)
LEARNING
Sampling
A subset of vehicles (observed)
Interactions between vehicle agents and link agents
Multi-agent Imitation Learning
Reproducing real-world traffic performance measures
MAGAIL-VL
SIMULATION
A subset of vehicles (generated)
Link traffic states (generated)
Network performance measures (route distribution, travel time, etc.)

2. Methodology

Imitation learning (IL) is a powerful alternative to Reinforcement Learning (RL) for learning sequential decision-making policies when manually defining *reward* functions is challenging. It attempts to recover an optimised reward function that could rationalise the expert demonstrations observed in the real data (Ho and Ermon, 2016). MAIL is an extension of IL which could learn multiple parametrized policies that imitate the behaviour of multiple experts from demonstrations of a set of experts interacting with each other in the same environment. MAGAIL is a specific algorithm of MAIL which applies generative adversarial networks (GAN) in the MAIL and includes a generator and a discriminator (Song et al., 2018). The generator controls the policies of all the agents, and the discriminator is a classifier trained to distinguish agent's behaviour from that of the corresponding expert.

2.1. Problem formulation

In MAGAIL-VL, a traffic network is formulated as a multi-agent system consisting of two groups of homogenous agents—links and (a subset of) vehicles—and their interactions are modelled using a Markov game containing N agents including n link agents $(1, \dots, i, \dots, n)$ and m vehicle agents $(n+1, \dots, j, \dots, n+m=N)$. The state of link agent i is: $s_i = \{i, linkstate\}$, where $linkstate \in [1, \dots, k]$ indicates the congestion level of link i (speed range). The action space for link agents represents possible changes in link congestion level: $\{A_i\}_{i=1}^{n} = A_l = \{+1, \dots, +(k-1), -1, \dots, -(k-1)\}$, which allows the state to vary between $\{i, 1\}, \dots, \{i, k\}$. The state of vehicle agent j is: $s_j = \{k, linkstate\}$, where k denotes the link that vehicle j is travelling on, and $linkstate$ is the congestion level of link k. The action space for vehicle agents represents possible movement of a vehicle on a link: $\{A_j\}_{j=n+1}^{N} = A_v = \{$ *Transfer to the leftmost link, the second leftmost link,…, the rightmost link, Stay on the same link, Enter the network, Exist the network, Stay outside the network before entering, Stay outside after* *existing*$\}$, where a vehicle can choose to move to a downstream link or stay at the current link at each time step. The function $\eta \in P(s)$ specifies the distribution of the initial states. Given that the agents are in state $\mathbf{s}^t$ at time t and agents take actions $(a_1, \ldots, a_N)$, the state changes to $\mathbf{s}^{t+1}$ with probability $P(\mathbf{s}^{t+1}|\mathbf{s}^t, a_1, \ldots, a_N)$. Each agent achieves its objective by selecting actions through a stochastic policy $\pi_i: S \to P(A_i)$. While different policies are to be specified for distinct agents in the original MAGAIL, to enables the model to incorporate a much larger number of agents in one multi-agent system, we define the same policy for link agents as π_l and the same policy for vehicle agents as π_v. The reward function of each agent is $r_i: S \times A_l^n \times A_v^m \to \mathbb{R}$. The reward function of agents: $\{r_i\}_{i=1}^n = r_l, \{r_j\}_{j=n+1}^N = r_v$. The goal of each agent is to maximise the total expected return $R_i = \sum_{t=0}^{\infty} \gamma^t r_{i,t}$, where γ is discount factor. The joint policy is defined as $\boldsymbol{\pi}(\boldsymbol{a}|\boldsymbol{s}) = \prod_{i=1}^n \pi_i(a_i|s_i) \prod_{j=n+1}^N \pi_j(a_j|s_i)$. The objective of this Markov game problem is to find the optimal reward functions and policies from the expert trajectories (vehicle trajectory and link state data) that could explain the expert behaviour (vehicle mobility pattern).

2.2. MAGAIL-VL

Using the GAN framework, MAGAIL-VL consists of the generator (G) to make realistic vehicle trajectories and link state changes based on the policies and the adversarial discriminator (D) to give reward feedback to the vehicle trajectories and link states generated by the generator until convergence. The policy and discriminator are both neural networks and the training process of MAGAIL-VL is as follows: With the initiated Markov game, we first generate vehicle trajectories and link states by rolling out the policies π_l and π_v for specific time steps. After sampling state-action pairs χ_E and χ_π from both expert trajectories and generated trajectories, respectively, the discriminator parameter could be updated by optimising the objective $\min \max \mathbb{E}_{\chi_\pi}\left[\sum_{i=1}^N \log\left(D_{\omega_i}(s, a_i)\right)\right] + \mathbb{E}_{\chi_E}\left[\sum_{i=1}^N \log\left(1 - D_{\omega_i}(s, a_i)\right)\right]$, where ω_i is the parameter set of the discriminator. Implicitly, D_{ω_i} plays the role of a reward function for the generator. Then the policy parameters are updated through reinforcement learning where a state-of-the-art natural policy gradient algorithm Multi-agent Actor-Critic with Kronecker-factors (Song et al., 2018) is used. The learned policies and reward functions are then obtained with repetition of this process. With the well-trained models, we can then generate vehicle trajectories and link states to simulate the network operation.

2.3. Baseline models

In addition to the proposed MAGAIL-VL model, we have developed several models for comparison. The first baseline model is MAGAIL-V which only considers the vehicle agents, while other configurations are similar to the MAGAIL-VL model. Additionally, we develop a single agent model base on GAIL (Ho and Ermon, 2016), while vehicle trajectories are generated sequentially by applying the model for specific times and no interaction exists between the vehicles. Moreover, we adopt the long short-term memory (LSTM) model and LSTM combined with self-attention mechanism (LSTM-attention) model as baseline models since they were demonstrated as efficient models in learning long-range location relations and predicting vehicle trajectories (Sun and Kim, 2021).

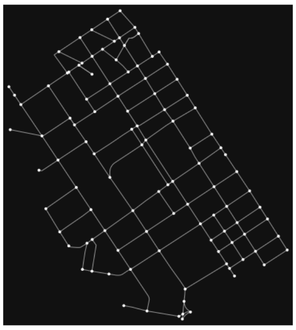
2.4. Evaluation measures

To evaluate the performance of proposed models, we employ *BLEU* (bilingual evaluation understudy) score to evaluate the accuracy of individual generated trajectories, where *BLEU-N* measures how consistent a model can generate *N* consecutive locations with an observed trajectory. For the network-level performance measures, we use the Jensen-Shannon distance of route probability distribution between generated and real trajectories, the mean absolute percentage error (MAPE) of average link travel times, and the mean absolute error (MAE) of link traffic states between the simulated network and the real network.

3. Data preparation

The data used in this study are extracted from the open traffic drone data collected in Athens, Greece through 20 datasets covering a few hours over four days (Barmpounakis and Geroliminis, 2020). We extract the data for a moderate-scale network (including 143 road links) as shown in Figure 2 with the time step of 10s. After map-matching vehicles' coordinate data using hidden Markov model (Meert and Verbeke, 2018), we obtained 600-800 vehicle trajectories for each dataset as population trajectories, where each trajectory dataset covers 80 time steps (800s). We then randomly sample 200 vehicle trajectories from each of the 20 datasets, resulting in 4000 vehicle trajectories in total. A sample of 200 vehicles per dataset represents approximately 25-33% of the population. The link states are identified according to the average speeds of vehicles on that link. We cluster five and three link states using K-means clustering method, for MAGAIL-VL-5 and MAGAIL-VL-3 model, respectively. In this study, we use the trajectories in first three days (15 datasets, training dataset) to train the models and test the models on the training dataset and further validate the model on the trajectories in the last day (5 datasets, validation dataset).

Figure 2: The study network



4. Results and Discussion

Based on the proposed models, we generated the trajectories of new 200 vehicles and link states for a simulation period of 800s, which corresponds to the observation period of the actual trajectory datasets. Note that the departure times and locations of vehicles are given as per the observed trajectories, and vehicles then travel to other locations based on the generator policy of models. We assess the model performance by comparing the consistency of generated dataset with the observed dataset with the proposed evaluation measures. The evaluation results for the training dataset and the validation dataset are provided in Figure 3 and Figure 4, respectively. The MAGAIL models perform better than other models in terms of the vehicle trajectory accuracy indicated by higher *BLEU* scores, demonstrating the effectiveness of capturing agent interactions in learning realistic vehicle movements. For the network-wide performance measures, the proposed MAGAIL-VL models overall outperforms other models, especially in link travel time measure. The performance of MAGAIL-V is quite poor in link travel time measure, suggesting the importance of considering both vehicle and link agents, rather than vehicle agents only. The traffic state measure results are reported only for the two MAGAIL-VL models with link agents that produce link state changes. The models with five traffic states and three traffic states have a MAE around 1 and 0.5, respectively, which indicates that the predicted link congestion level is on average half-state different from the actual link congestion level. Overall, MAGAIL-VL provides the most satisfactory performance in both trajectory-level and network-wide measures. This study demonstrates the possibility of modelling the network-wide traffic state evolution by learning the interaction between only a subset of vehicles and the surrounding link congestion levels. This finding provides an important first step towards enabling a fully data-driven traffic simulation model for a large-scale network in a more effective and efficient manner.

Figure 3: Evaluation results for the training dataset

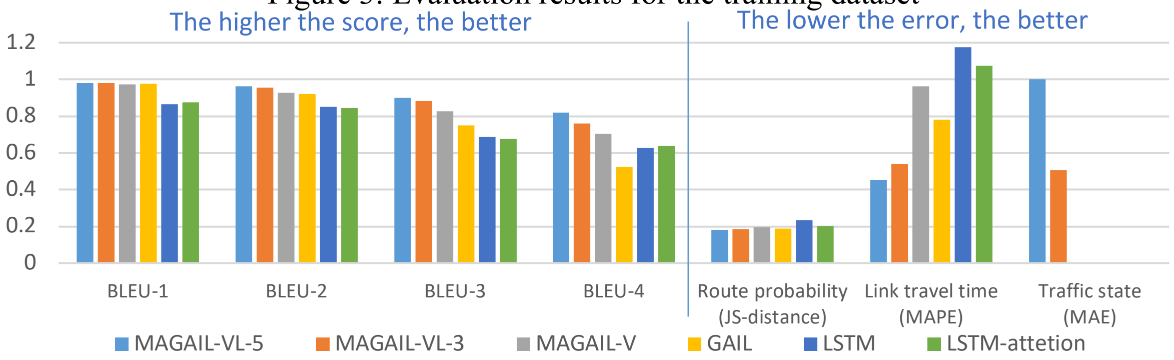
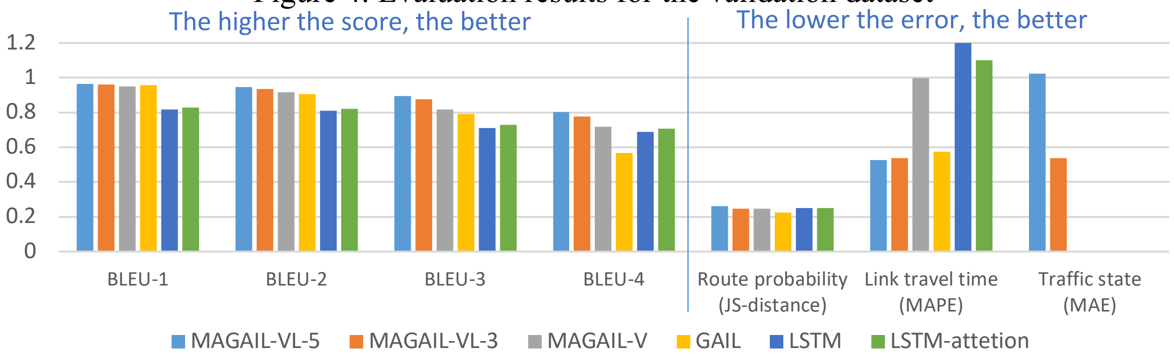


Figure 4: Evaluation results for the validation dataset



Acknowledgements

This research was partially funded by the Australian Research Council under grant DE190101020.

References

BARMPOUNAKIS, E. & GEROLIMINIS, N. 2020. On the new era of urban traffic monitoring with massive drone data: The pNEUMA large-scale field experiment. *Transportation research part C: emerging technologies,* 111**,** 50-71.

BHATTACHARYYA, R. P., PHILLIPS, D. J., WULFE, B., MORTON, J., KUEFLER, A. & KOCHENDERFER, M. J. Multi-agent imitation learning for driving simulation. 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2018. IEEE, 1534-1539.

CHOI, S., KIM, J. & YEO, H. 2021. TrajGAIL: Generating urban vehicle trajectories using generative adversarial imitation learning. *Transportation Research Part C: Emerging Technologies,* 128**,** 103091.

HO, J. & ERMON, S. 2016. Generative adversarial imitation learning. *Advances in neural information processing systems,* 29**,** 4565-4573.

KIM, J. & MAHMASSANI, H. S. 2015. Spatial and temporal characterization of travel patterns in a traffic network using vehicle trajectories. *Transportation Research Part C: Emerging Technologies,* 59**,** 375-390.

LI, L., JIANG, R., HE, Z., CHEN, X. M. & ZHOU, X. 2020. Trajectory data-based traffic flow studies: A revisit. *Transportation Research Part C: Emerging Technologies,* 114**,** 225-240.

MAHMASSANI, H. S. 2001. Dynamic network traffic assignment and simulation methodology for advanced system management applications. *Networks and spatial economics,* 1**,** 267-292.

MEERT, W. & VERBEKE, M. HMM with non-emitting states for Map Matching. European Conference on Data Analysis (ECDA), Date: 2018/07/04-2018/07/06, Location: Paderborn, Germany, 2018.

SONG, J., REN, H., SADIGH, D. & ERMON, S. 2018. Multi-agent generative adversarial imitation learning. *arXiv preprint arXiv:1807.09936*.

SUN, J. & KIM, J. 2021. Joint prediction of next location and travel time from urban vehicle trajectories using long short-term memory neural networks. *Transportation Research Part C: Emerging Technologies,* 128**,** 103114.

ZIEBART, B. D., MAAS, A. L., BAGNELL, J. A. & DEY, A. K. Maximum entropy inverse reinforcement learning. AAAI, 2008. Chicago, IL, USA, 1433-1438.