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

Car Following Modelling with Constrained Generative Adversarial Imitation Learning

Lin Lin[1], Jiwon Kim[1], Sanghyung Ahn[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

Traffic simulation plays an important role in traffic planning and management. Since microscopic modeling can capture individual vehicle movements and interactions at the highest level of detail, it is implemented in major traffic simulation software and car following is a significant component in those simulators. Most of the existing car following models are theory- or model-based in that a driver's action (e.g., acceleration) is described as an analytical function of a set of parameters representing traffic states and driver characteristics, which are based on physics or behavioural theories. While various car-following models have been widely adopted in microscopic simulation research, such as Gipps model and Intelligent Driver Model (IDM), there are limitations of these theory-based models, one of which being the limited number of parameters that are often insufficient to capture and generalise complex human behaviours. On the other hand, data-driven, machine learning models like neural networks can embed and generalise much richer information through a more flexible model structure and larger parameter set. As such, there is an increasing interest in developing data-driven traffic simulation models in the transport research community and this study aims to investigate the possibility of developing a car-following model using emerging deep learning techniques.

Among deep learning models, Recurrent Neural Network (RNN) has been widely used in solving sequential problems due to its ability to learn long-term dependencies on sequential data. In 2014, Generative Adversarial Network (GAN) [1] was introduced, which has become the most popular algorithm used in synthetic data generation including image, audio and text generation. Several studies that combine Long Short-Term Memory (LSTM) network, which is a kind of RNN, and GAN obtain novel results on synthetic trajectory generation [2]. Meanwhile, Reinforcement Learning (RL), which is formulated based on Markov Decision Process (MDP) and aims to solve sequential decision-making problems, saw significant improvement after integrating with deep neural networks and produced well-known applications like AlphaGo and self-driving cars. In RL, the key is to define a reward function that gives an RL agent proper signals to guide its actions. However, it is often difficult to manually design the reward function in complex problems and, thus, Inverse Reinforcement Learning (IRL) [3] and Imitation Learning (IL) were proposed, which can learn the reward function from experts' demonstrations observed from data. In 2016, Generative Adversarial Imitation Learning (GAIL) [4] was proposed, which is a combination of IL and GAN. In GAIL, training is conducted through a zero-sum competition between a policy network and a discrimination network, where the discriminator serves as a reward signal for the RL problem. Recently, studies have applied GAIL in the context of car-following modelling [5][6]. However, due to the nature of IL that learns a reward function solely from data, there is little control over the undesirable behaviours of the RL agent representing a driver. For instance, the driver may end up following too close to the leading car, resulting in a collision, or even moving backward as there are no explicit signals to avoid such states within the IL framework. To address this challenge, we propose a *Constrained-GAIL* framework that utilises the *reward augmentation* technique that was introduced in a multi-agent traffic study [7]. This reward augmentation allows additional constraints to be imposed through a manually designed reward function, on top of the reward function learned from the original GAIL. In this way, driving agents can be guided to avoid unwanted states, thereby reducing unrealistic or dangerous events. We demonstrate our proposed model using the NGSIM dataset, and the result shows higher prediction accuracy in terms of vehicle speed and location, while minimizing unwanted behaviours including collision and backward driving.

2. Method

2.1. Formulation

The car-following behaviour of a driver is formulated as an MDP in the RL setting, where the RL agent represents the following car and aims to learn the optimal policy (π) through trial and error. A policy is a stochastic rule by which the agent selects actions as a function of states and the agent keeps correcting its policy by comparing its trajectories with the ground-truth trajectories (expert demonstrations) from the dataset.

The MDP is defined as a tuple (S, A, T, R, γ), where S is the state space including the speed of the following car, speed difference and distance to the leading car, A is the continuous action space representing the range of acceleration values the following car can choose, T stands for the transition probabilities describing the probability of moving one state to another given an action, R denotes the reward function obtained from the demonstrated trajectories in the dataset, and γ denotes the discount factor reflecting how much the agent cares about the reward in the distant future compared to the immediate reward. At every time-step, the RL agent (following car) observes a state and chooses an acceleration value based on its policy (π). It then gets to the next state according to the transition probabilities, meanwhile receiving a reward. In most cases, the transition probabilities are unknown, but the environment is fully observable to get the next state information. The aim to solve the MDP is to get the optimal policy that can give the largest accumulated reward in a complete journey from the origin to the destination.

In GAIL, a discriminator is introduced to the above-mentioned RL context and tries to distinguish state-action pairs of the trajectories generated by the agent's policy (π) from those generated by the expert's policy (π_e). Learning becomes a *minimax game* between the discriminator and the policy, where the discriminator tries to maximize its classification ability and policy tries to generate realistic trajectories to fool the discriminator, as shown in following formulation:

$$\min_{\pi} \max_{D\in(0,1)^{S\times A}} \boldsymbol{E}_{\pi}[\log D(s,a)] + \boldsymbol{E}_{\pi_e}[\log(1-D(s,a))] \; - \lambda H(\pi)$$

where π is the policy imitating expert policy π_e, D is the discriminator, and H(π) is the causal entropy of the policy π. The goal is to find policy π that minimises the distance between the distribution of generated state-action pairs and the distribution of expert state-action pairs. Since the discriminator gives higher scores to ground-truth state-action pairs and lower scores to generated ones, it can be served as a reward signal in the RL problem. The reward function is thus obtained from expert demonstrations directly through discriminator.

The constrained-GAIL minimax problem is solved by transforming to an unconstrained form of GAIL to a constrained form by adding a regularizer to the formulation as follows:

$$\min_{\pi_\theta} \max_{D\in(0,1)^{S\times A}} \boldsymbol{E}_{\pi_\theta}[\log D_\omega(s,a)] + \boldsymbol{E}_{\pi_e}[\log(1-D_\omega(s,a))] \; - \lambda H(\pi) - rE_\pi[I_u]$$

where r is the penalty and I_u is an indicator function which will be zero if the state-action pair does not lead to an undesired state [7]. The policy and discriminator are represented by deep neural networks parameterized by θ and ω. The constrained GAIL model is trained to find the optimal parameters by alternating between a gradient step to increase the formula above with respect to the discriminator parameters ω and a Proximal Policy Optimization (PPO) step to decrease above formula with respect to θ. The optimal policy and discriminator will be obtained after training is complete.

2.2. Model Structure

The model architecture of the proposed constrained GAIL to predict a driver's car-following behaviour is shown in Figure 1. At each time step, the generator outputs the following car's acceleration based on the input state considering the current speed and the relation with the leading car. These generated state-action pairs along the whole trajectory are then compared with the state-action pairs from the real dataset through the discriminator. Both the generator network and the discriminator network got optimized through the training iterations.

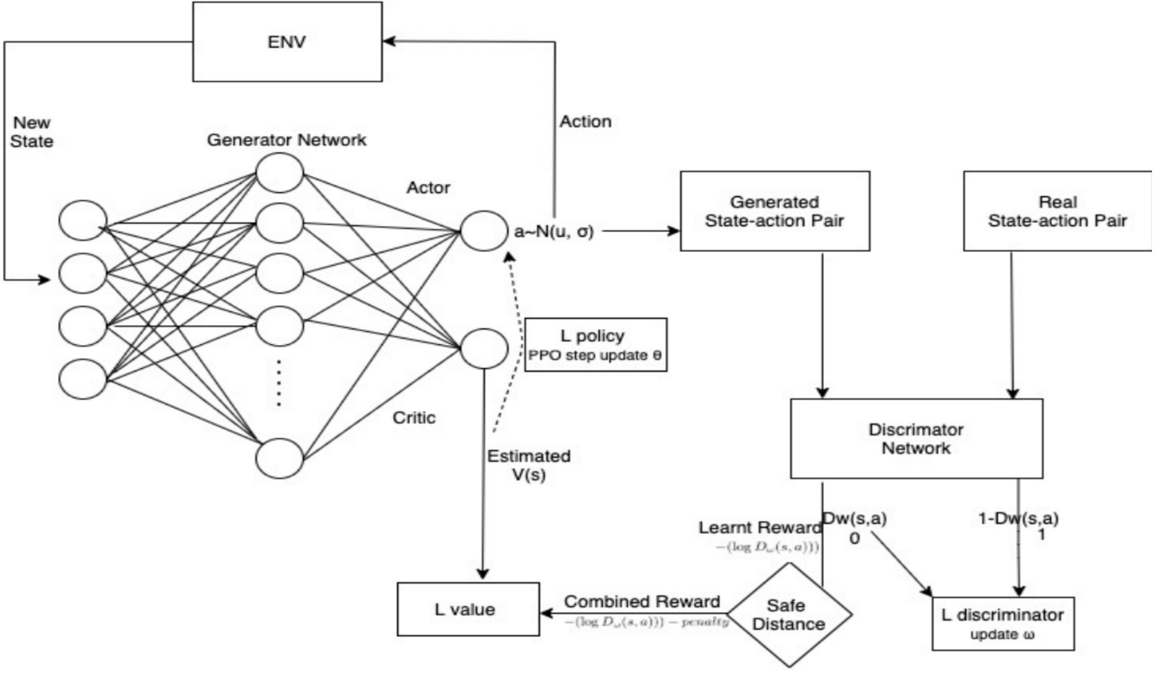


Figure 1: Network structure

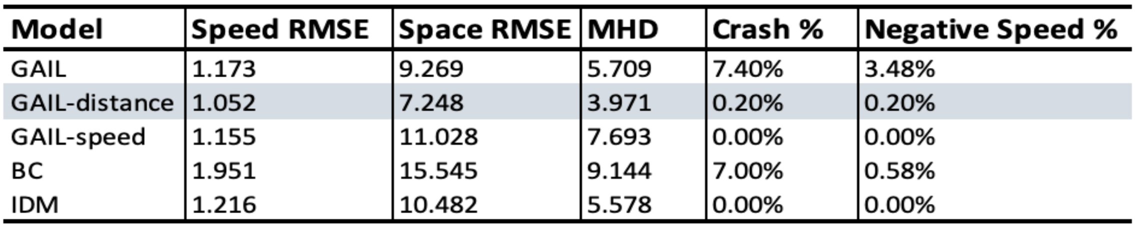
The generator network utilizes an *Actor-Critic* framework, and the PPO algorithm is used to update the gradient of the policy. With each update, the new policy should have a higher probability of generating actions with larger rewards than the previous policy. In the Actor component, a distribution of actions is outputted and the acceleration of the following car is actually drawn from this distribution, while from the Critic component a value evaluating the goodness of the action is outputted. The discriminator network is in charge of judging how close the real state-action pairs from the dataset and the generated state-action pairs from the policy network are. The cross-entropy of these two sets of state-action pairs is calculated and the training is conducted to find the parameters that lead to the maximum cross-entropy. With the extra augmented reward, rules or constraints can be integrated into the network. In this study, two types of constraints are considered: one specifying the minimum distance between the leader and follower to prevent a collision and the other not allowing a negative speed to prevent driving backward. A constant penalty is added if the distance between the following car and the leading car is less than 2 meters or when the speed is negative.

2.3. Dataset

This work uses the NGSIM I-80 dataset [8], which contains 45 minutes of detailed trajectory data from eastbound I-80 in the San Francisco Bay area in Emeryville, CA, on April 13, 2005. A total of 889 pairs of car-following trajectories are extracted from the dataset, where vehicles with lane change behaviours have been excluded. Among them, 647 pairs (4:00 - 4:15 p.m.; 5:00 - 5:15 p.m.) are used as training data and 242 pairs (5:15 p.m. - 5:30 p.m.) as testing data.

3. Result

The accuracy of the constrained GAIL model is measured by the Root Mean Square Error (RMSE) of speed and location and the Modified Hausdorff Distance (MHD) [9] of location between the modelled trajectories and the ground-truth trajectories. The ability to prevent unwanted behaviours is measured by the frequencies of collisions (collision ratio) and backward driving observations (negative speed ratio). The model performance was measured for the distance-constrained version (GAIL-distance) and the speed-constrained version (GAIL-speed), separately. These two models are also compared with other benchmark models including the original GAIL model, IDM, and Behavior Cloning (BC) model. Table 1 shows the performance comparison results. The distance-constrained GAIL achieved the lowest error among all the models, where collision and backward driving frequencies are much lower in GAIL-distance than in the original GAIL and BC model. IDM achieves the similar speed RMSE with GAIL, but shows higher location RMSE.



Model	Speed RMSE	Space RMSE	MHD	Crash %	Negative Speed %
GAIL	1.173	9.269	5.709	7.40%	3.48%
GAIL-distance	1.052	7.248	3.971	0.20%	0.20%
GAIL-speed	1.155	11.028	7.693	0.00%	0.00%
BC	1.951	15.545	9.144	7.00%	0.58%
IDM	1.216	10.482	5.578	0.00%	0.00%

Table 1: The results of Speed RMSE, Location RMSE, MHD, Collision Ratio and Negative Speed Ratio

The spacing (the front-to-front distance) between leading and following cars generated by different models are shown in Figure 2 to investigate the effectiveness of adding extra constraints through reward augmentation to the GAIL framework. For a given leading car trajectory, the spacing results produced by the original GAIL, the proposed distance-constrained GAIL (GAIL-distance), and IDM model are compared to the ground-truth results (NGSIM). Abnormal overtakings occur in the original GAIL, which are reflected by the negative spacing values, while such abnormal behaviours are not observed in the distance-constrained GAIL, indicating the reward augmentation successfully imposes the desired constraints.

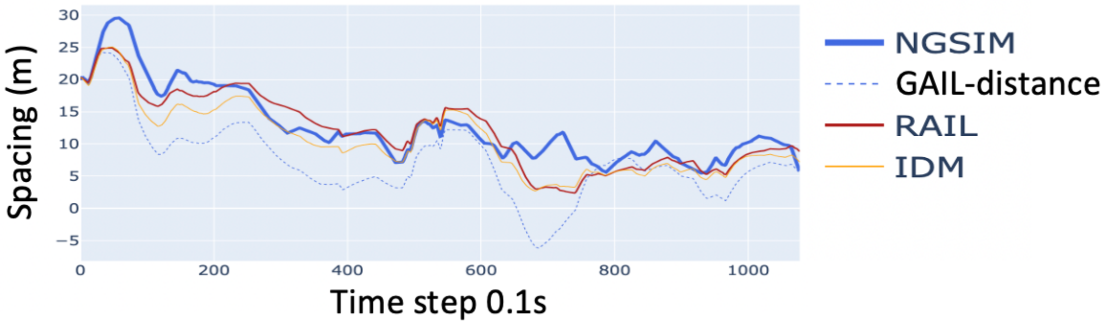


Figure 2: Spacing between following car (#1543) and leading car (#1539) from the NGSIM data and the models

The similarity between the model-generated trajectories and the NGSIM trajectories is also evaluated by comparing the speed distributions, as presented in Figure 3(a). The distance-constrained GAIL eliminates negative speeds and produces the speed distribution similar to the actual NGSIM data, whereas the original GAIL produces negative speeds. Compared to IDM, the distance-constrained GAIL matches the speed distribution of the NGSIM better, especially in the right tail where similar high speed ranges are observed between the GAIL-distance and NGSIM, as shown in Figure 3(b).

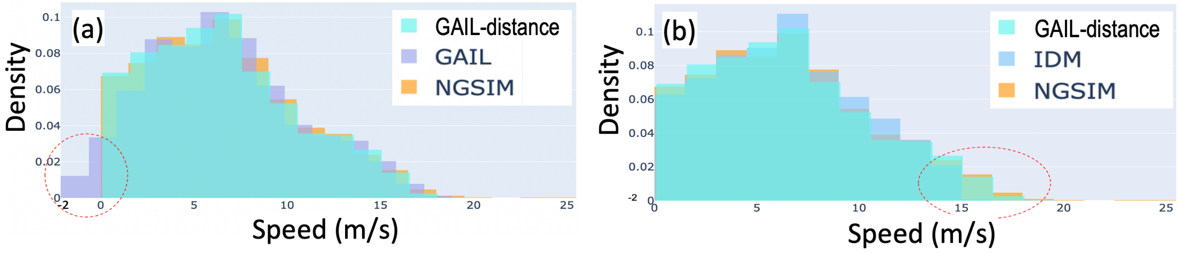


Figure 3: The comparison of speed distributions among (a) GAIL-distance, GAIL, and NGSIM data (b) GAIL-distance, IDM, and NGSIM data

4. Conclusion

This study demonstrates a data-driven car-following model based on GAIL, where augmented reward signals can be integrated into the imitation learning framework to impose extra constraints or penalties to more effectively guide a learning agent's behaviour. The experiments conducted using the NGSIM trajectory data show that the proposed constrained-GAIL framework can improve the model prediction accuracy, compared to the conventional theory-based car-following model (IDM) and the existing imitation learning models (GAIL and BC), while successfully preventing unwanted behaviours like collision or backward driving, which is crucial to perform a realistic simulation of safety-critical scenarios. Several extensions and improvements to the current model are possible. Since the proposed framework is general and flexible, it could be applied to other microscopic traffic simulation problems like lane changing or multi-agent scenarios. Another improvement could be extending the models with latent variables like infoGAIL, which can capture and classify different behaviours and patterns in an unsupervised manner.

Acknowledgements

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

References

[1] I. J. Goodfellow *et al.*, "Generative Adversarial Networks," *arXiv:1406.2661 [cs, stat]*, Jun. 2014.
[2] J. Rao, S. Gao, Y. Kang, and Q. Huang, "LSTM-TrajGAN: A Deep Learning Approach to Trajectory Privacy Protection," *arXiv:2006.10521 [cs]*, Jun. 2020.
[3] Abbeel, P.; Ng, A.Y. Apprenticeship learning via inverse reinforcement learning. In Proceedings of the Twenty-First International Conference on Machine Learning, Banff, AL, Canada, 4–8 July 2004; p. 1.
[4] J. Ho and S. Ermon, "Generative Adversarial Imitation Learning," *arXiv:1606.03476 [cs]*, Jun. 2016.
[5] Zhou, Y., Fu, R., Wang, C., Zhang, R. (2020). Modeling Car-Following Behaviors and Driving Styles with Generative Adversarial Imitation Learning. Sensors, 20(18), 5034.
[6] G. Zheng, H. Liu, K. Xu, and Z. Li, "Learning to Simulate Vehicle Trajectories from Demonstrations," in 2020 IEEE 36th International Conference on Data Engineering (ICDE), Dallas, TX, USA, Apr. 2020, pp. 1822–1825. doi: 10.1109/ICDE48307.2020.00179.
[7] R. P. Bhattacharyya, D. J. Phillips, C. Liu, J. K. Gupta, K. Driggs-Campbell, and M. J. Kochenderfer, "Simulating Emergent Properties of Human Driving Behavior Using Multi-Agent Reward Augmented Imitation Learning," arXiv:1903.05766 [cs], Mar. 2019.
[8] Next Generation Simulation (NGSIM) datasets, https://ops.fhwa.dot.gov/trafficanalysistools/ngsim.htm.
[9] Zhou, Y., Fu, R., & Wang, C. (2020). Learning the car-following behavior of drivers using maximum entropy deep inverse reinforcement learning. Journal of Advanced Transportation, 2020, 1–13.