2022 IEEE/ACM 7th Symposium on Edge Computing (SEC) | 978-1-6654-8611-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/SEC54971.2022.00043

Poster: Reliable On-Ramp Merging via Multimodal Reinforcement Learning

Gaurav Bagwe[1], Jian Li[2], Xiaoheng Deng[3], Xiaoyong Yuan[4], Lan Zhang[1]
[1]Department of Electrical and Computer Engineering, Michigan Technological University, Houghton, MI, USA
[2]School of Cyber Science and Technology, University of Science and Technology of China, Hefei, China
[3]School of Computer Science and Engineering, Central South University, Changsha, China
[4]College of Computing, Michigan Technological University, Houghton, MI, USA
{grbagwe, xyyuan, lanzhang}@mtu.edu, lijian9@ustc.edu.cn, dxh@csu.edu.cn

***Abstract*—The recent success of Artificial Intelligence (AI) has enabled autonomous driving with better perception capabilities. However, on-ramp merging remains one of the main challenging scenarios for reliable autonomous driving. Within the limited onboard sensing range, a merging vehicle can hardly observe and predict the main road conditions properly, restricting appropriate merging maneuvers. In this poster, we outline ongoing research ideas for reliable and autonomous on-ramp merging assisted by vehicular communications. By jointly leveraging the basic safety messages (BSM) from neighboring vehicles and the surveillance images, a merging vehicle can perform reliable driving via robust multimodal reinforcement learning. Some experimental results are provided to evaluate our idea under the Simulation of Urban MObility (SUMO) platform.**

I. Introduction

On-ramp merging has been the main bottleneck of freeway driving. Due to the reduced traffic capacity and various driving behaviors, improper handling and merging may lead to severe traffic congestion and even accidents [1]. Although automakers and researchers have developed and tested different levels of automated driving safety designs, such as the advanced driver assistance systems (ADAS) [2], these systems mainly rely on onboard sensors with a limited sensing range. Hence, a merging vehicle relying on such onboard sensors can hardly observe the main road conditions automatically. With a larger communication range, connected and automated vehicles (CAVs) have the potential to observe main road conditions by communicating with vehicles on the main road [3], [4]. However, a merging vehicle needs to properly change its speed, select a target merging position on the main road, and smoothly change the lane to merge. Moreover, a merging vehicle not only considers its own driving state but also cooperates with surrounding vehicles, predicts their driving behaviors, and finally determines the merging maneuvers, requiring very high sensing and path planning ability. Therefore, in this poster, we present our ongoing research ideas for reliable and robust on-ramp merging based on reinforcement learning (RL), named Robust Augmented Multi-modal and Reinforcement Learning (RAMRL), as shown in Fig. 1.

Specifically, in addition to the basic safety messages (BSM) shared by nearby vehicles or roadside units (RSU) [3], a CAV uses images captured from the roadside surveillance camera to properly perform on-ramp merging. We use RL to automatically learn from environmental observations under supervision based on the environment's reward returns [5]. The merging problem is formulated as a Markov decision process (MDP) by jointly considering safety, comfort, and traffic efficiency for on-ramp merging. We implement one popular model-free RL algorithm, the proximal policy optimization (PPO) network, because of its higher performance in continuous high-dimensional action spaces [6]. Some experimental results are provided to evaluate our idea under the Simulation of Urban MObility (SUMO) platform [7].

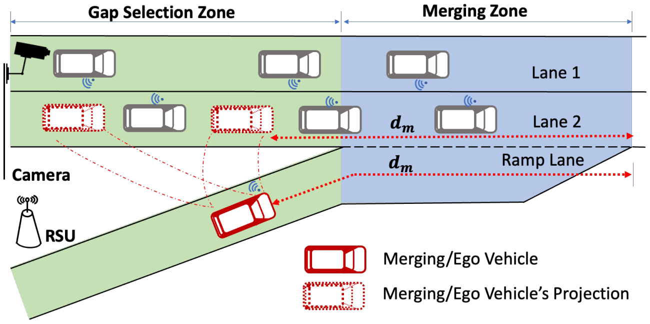


Fig. 1: On-Ramp Merging Problem : General representation of merging section where the ego vehicle (in red) needs to navigate from the ramp lane to lane 2 on the highway. The scenario consists of the roadside unit which communicates with all the vehicles on the road and a camera which provides the input image of the merging section.

II. Proposed Design

We consider a general on-ramp merging problem. Taking the merging topology in Fig. 1 as an example[1], a highway section consists of lane 1, lane 2, and a parallel-type ramp, where a part of the ramp is parallel to lanes 1 and 2 [1]. Each CAV has communication and computation capabilities and is equipped with an On-Board Unit (OBU) that can communicate to the RSU [8]. Besides, a surveillance camera is deployed at the roadside, as shown in Fig. 1, which captures driving images to the RSU. Consider a vehicle that is trying to merge with the traffic on the highway using the ramp, named by a merging

[1] Our approach can be generalized to different merging topologies.

978-1-6654-8611-8/22/$31.00 ©2022 IEEE
DOI 10.1109/SEC54971.2022.00043

vehicle or ego vehicle. Our design aims to develop an efficient merging strategy for the ego vehicle based on multi-modal merging knowledge from the camera and RSU. We formulate this merging problem as a Markov Decision Process (MDP). At each time step t, the agent, *i.e.*, the ego vehicle, observes a state and takes action. This action is implemented in the environment, which returns the next state and a reward. We describe main contents of the formulated MDP problem below.

- *State Space*: An ego vehicle needs information on its own state and the states of other vehicles on the highway [9]. In our multimodality design, the state information from the BSM consists of the current location and velocity of the vehicles on the road [10], [11]; the state information from the surveillance image captures the traffic conditions that inherently incorporate the vehicles' driving behaviors and the highway traffic topology.
- *Action Space*: Based on the current position of an ego vehicle, the action space is different. As shown in Fig. 1, when the ego vehicle is located at the gap selection zone, the action space is the acceleration space; when it is located in the merging zone, the action space consists of both the acceleration and the steering angle.
- *Reward*: Reward is a type of feedback given to the agent to inform if the action is favorable. Similar to the action space design, a two-stage reward is developed based on the ego vehicle's position. We intend to jointly optimize merging safety, comfort, and efficiency. Due to the page limit, we only provide the safety measurement r_d:

$$r_d = 2 - (\frac{d_f}{d_{max}})^{-\alpha_f} + (\frac{d_r}{d_{max}})^{-\alpha_r}, \tag{1}$$

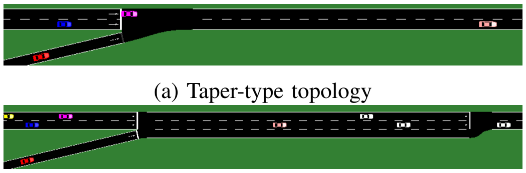
  where d_{max} is the distance of ramp; d_f is the distance between the projected ego vehicle and the first following vehicle; d_r is the distance between the projected ego vehicle and the first preceding vehicle; $\alpha \in [0, 1]$ is the corresponding hyperparameter.

We use the model-free policy-based RL algorithms to solve the above MDP problem, as they are suitable for continuous action spaces and have better convergence [12], [13]. Schulman *et al.* introduced a model-free policy optimization method called Trust Region Policy Optimization (TRPO) [14]. TRPO updates the policy with a constraint on the size of the update [6]. The surrogate objective function is defined as:

$$\underset{\theta}{maximize}\ L_{\theta_{old}}(\theta) \tag{2}$$

$$s.t.\ D_{KL}^{max}(\theta_{old}, \theta)) \leq \delta, \tag{3}$$

where θ is the network parameters for the policy $\pi_\theta(a|s)$), and θ_{old} are the parameters of the old policy. This policy is updated on the trust region constraint of KL divergence denoted by D_{KL}, where δ is the bound on KL divergence [14]. The proximal policy optimization (PPO) algorithm uses a first-order derivative, which makes it as reliable and efficient as the TRPO while being simpler to implement [6]. Specifically, PPO uses a clipped surrogate objective which is a modification of the objective of the TRPO network. The network is trained to maximize the expected returns from the loss function [6], which can be given by


(a) Taper-type topology

(b) Parallel-type topology

Fig. 2: (a) Taper-type topology: the ego vehicle does not have an acceleration lane before the merge intersection. (b) Parallel-type topology: the ego vehicle has an additional parallel lane.

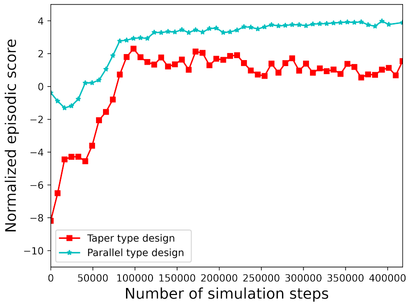


Fig. 3: Training curves for two type merging scenarios.

$$\begin{aligned} &L_t^{CLIP+VF+S}(\theta) = \\ &\hat{\mathbb{E}}_t\left[L_t^{CLIP}(\theta) - c_1 L_t^{VF}(\theta) + c_2 S_{[\pi_\theta(s_t)]}\right], \end{aligned} \tag{4}$$

where c_1 and c_2 are coefficients and S denotes the entropy bonus. L_t^{CLIP} is the clipped objective, $L_t^{VF}(\theta)$ is the Mean Squared error between the value function and the target value function, and the third term gives entropy bonus.

III. Experimental Results

We develop an on-ramp merging simulator using the Simulation of Urban MObility (SUMO) platform [7]. SUMO is an open-source microscopic traffic simulator consisting of rule-based car-following models to simulate real-world driving conditions. These car-following models can be controlled online using Traffic Control Interface (TraCI). TraCI is an interface between SUMO and python, which allows information retrieval and manipulations of objects such as vehicles, and traffic lights [15]. To create the environment, we build on the sumo-rl environment [16], which is primarily built for traffic signal control. We illustrate the training performance of our design under two merging types in Fig. 2 and Fig. 3. We observe that the taper type design earlier to train for the RL algorithm, which validates the merge type design preferred by AASHTO [17]. Higher episodic rewards in the parallel type design ensure driving comfort, safety, and lower merge time.

IV. Next Step

We will take safety, comfort, and efficiency into account by formulating comprehensive MDP problems with advanced RL techniques for reliable and robust merging.

References

[1] F. J. Koepke, “Ramp exit/entrance design–taper versus parallel and critical dimensions,” *Transportation Research Record*, no. 1385, 1993.

[2] J. Borrego-Carazo, D. Castells-Rufas, E. Biempica, and J. Carrabina, “Resource-constrained machine learning for adas: a systematic review,” *IEEE Access*, vol. 8, pp. 40 573–40 598, 2020.

[3] P. Kopelias, E. Demiridi, K. Vogiatzis, A. Skabardonis, and V. Zafiropoulou, “Connected & autonomous vehicles–environmental impacts–a review,” *Science of the total environment*, vol. 712, p. 135237, 2020.

[4] L. Zhang, L. Yan, Y. Fang, X. Fang, and X. Huang, “A machine learning-based defensive alerting system against reckless driving in vehicular networks,” *IEEE Transactions on Vehicular Technology*, vol. 68, no. 12, pp. 12 227–12 238, 2019.

[5] W. Zhao, J. P. Queralta, and T. Westerlund, “Sim-to-real transfer in deep reinforcement learning for robotics: a survey,” in *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 2020, pp. 737–744.

[6] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, “Proximal policy optimization algorithms,” *arXiv preprint arXiv:1707.06347*, 2017.

[7] P. A. Lopez, M. Behrisch, L. Bieker-Walz, J. Erdmann, Y.-P. Flötteröd, R. Hilbrich, L. Lücken, J. Rummel, P. Wagner, and E. Wießner, “Microscopic traffic simulation using sumo,” in *2018 21st international conference on intelligent transportation systems (ITSC)*. IEEE, 2018, pp. 2575–2582.

[8] NHTSA. (2016) Federal motor vehicle safety standards; v2v communications. [Online]. Available: https://www.nhtsa.gov/sites/nhtsa.gov/files/documents/v2v_nprm_web_version.pdf

[9] I. Nishitani, H. Yang, R. Guo, S. Keshavamurthy, and K. Oguchi, “Deep merging: Vehicle merging controller based on deep reinforcement learning with embedding network,” in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 216–221.

[10] N. Nuthalapati, V. S. Koganti, and L. K. Galla, “Reliability of warning light device applications,” in *2016 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)*. IEEE, 2016, pp. 425–430.

[11] R. V. Cowlagi, R. C. Debski, and A. M. Wyglinski, “Risk quantification for automated driving using information from v2v basic safety messages,” in *2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring)*, 2021, pp. 1–5.

[12] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[13] H. D. Z. Ding, S. Zhang, H. Yuan, H. Zhang, J. Zhang, Y. Huang, T. Yu, H. Zhang, and R. Huang, *Deep Reinforcement Learning: Fundamentals, Research, and Applications*, S. Z. Hao Dong, Zihan Ding, Ed. Springer Nature, 2020, http://www.deepreinforcementlearningbook.org.

[14] J. Schulman, S. Levine, P. Abbeel, M. Jordan, and P. Moritz, “Trust region policy optimization,” in *International conference on machine learning*. PMLR, 2015, pp. 1889–1897.

[15] C. Wu, A. Kreidieh, K. Parvate, E. Vinitsky, and A. M. Bayen, “Flow: A modular learning framework for autonomy in traffic,” *arXiv preprint arXiv:1710.05465*, 2017.

[16] L. N. Alegre, “SUMO-RL,” https://github.com/LucasAlegre/sumo-rl, 2019.

[17] A. AASHTO, “Policy on geometric design of highways and streets,” *American Association of State Highway and Transportation Officials, Washington, DC*, vol. 1, no. 990, p. 158, 2001.