Learning a Robust Multiagent Driving Policy for Traffic Congestion Reduction

Yulin Zhang[1], William Macke[1], Jiaxun Cui[1], Daniel Urieli[2], and Peter Stone[1,3]

[1] The University of Texas at Austin, USA
[2] General Motors R&D Labs, Israel
[3] Sony AI, USA
{yulin,wmacke,pstone}@cs.utexas.edu; jiaxun@utexas.edu
daniel.urieli@gm.com

Abstract. In most modern cities, traffic congestion is one of the most salient societal challenges. Past research has shown that inserting a limited number of autonomous vehicles (AVs) within the traffic flow, with driving policies learned specifically for the purpose of reducing congestion, can significantly improve traffic conditions. However, to date these AV policies have generally been evaluated under the same limited conditions under which they were trained. On the other hand, to be considered for *practical* deployment, they must be robust to a wide variety of traffic conditions. This paper establishes for the first time that a multiagent driving policy can be trained in such a way that it generalizes to different traffic flows, AV penetration, and road geometries, including on multi-lane roads.

Keywords: Autonomous Vehicles · Traffic Optimization · Deep Reinforcement Learning· Multiagent Systems

1 Introduction

According to Texas A&M's 2021 Urban Mobility Report, traffic congestion in 2020 in the U.S. was responsible for excess fuel consumption of about 1.7 billion gallons, an annual delay of 4.3 billion hours, and a total cost of \$100B [3]. A common form of traffic congestion on highways is *stop-and-go waves*, which have been shown in field experiments to emerge when vehicle density exceeds a critical value [6]. Past research has shown that in human-driven traffic, a small fraction of automated or autonomous vehicles (AVs) executing a controlled multiagent driving policy can mitigate stop-and-go waves in simulated and real-world scenarios, roughly double the traffic speed, and increase throughput by about 16% [5]. Frequently, the highest-performing policies are those learned by deep reinforcement learning (DRL) algorithms, rather than hand-coded or model-based driving policies.

Any congestion reduction policy executed in the real world will need to perform robustly under a wide variety of traffic conditions such as traffic flow, AV penetration (percentage of AVs in traffic, referred to here as "AVP"), AV placement in traffic, and road geometry. However, existing driving policies have generally been tested in the same conditions they were trained on, and have not been thoroughly tested for robustness to different traffic conditions. Therefore, it remains unclear how to create a robust DRL congestion-reduction driving policy that is practical for real-world deployment.

Fig. 1: a single-lane merge road.

In this extended abstract (of full paper[7]), we establish for the first time the existence of a robust DRL congestion-reduction driving policy that performs well across a wide variety of traffic flows, AVP, AV placement in traffic, and several road geometries. Moreover, we investigate the question of how to come up with such a policy and what degree of robustness it can achieve. We create a benchmark with a diverse, pre-defined collection of test traffic conditions of real-world interest including the single-lane merge scenario shown in Figure 1. Such merge scenarios are a common source of stop-and-go waves on highways [4]. While there are different approaches to training robust DRL policies in other domains with different levels of success, our approach is to systematically search for a robust policy by varying the training conditions, evaluating the learned policy on our proposed test set in a single-lane merge scenario, and selecting the highest performing one. The highest performing policy outperforms the human-only baseline with as few as $1\,\%$ AVs across different traffic conditions in the single-lane merge scenario. We further investigate the policy's generalization to more complex roads it has not seen during training, specifically with two merging ramps at a variety of distances, or on a double-lane main road with cars able to change lanes. Notwithstanding negative prior results showing that a policy developed in a single-lane ring road fails to mitigate the congestion on a double-lane ring road [2], the learned policy outperforms human-only traffic and effectively mitigates congestion in all of these scenarios defined by our benchmark. Taken together, this paper's contributions and insights take us a step closer towards making the exciting concept of traffic congestion reduction through AV control a practical reality.

2 Robustness evaluation conditions and metrics in merge road

Similarly to past work [1], our baseline setup consists of simulated human-driven vehicles only. In contrast to past work, which typically showed improvement over this baseline in a *single* combination of traffic conditions, our goal is to develop a robust AV driving policy that improves over this baseline across a *range* of realistic traffic conditions in the merge road shown in Figure 1, characterized by:

- *Main Inflow Rate*: the amount of incoming traffic on the main artery (veh/hour),
- *Merge Inflow Rate*: the amount of incoming traffic on the merge road (veh/hour),
- *AV Placement*: the place where the AVs appear in the traffic flow; the AVs can either be distributed evenly or randomly among the simulated human-driven vehicles.
- *AV Penetration*: the percentage of vehicles that are controlled autonomously,
- *Merge road geometry:* the distance between two merge junctions (in relevant scenarios), and the number of lanes.

In this paper, we fix the merge inflow rate to be $200\,\mathrm{veh/hour}$ (small enough to cause traffic congestion on the main road) and set the range of the main inflow to be [1600, 2000] veh/hour (resulting in minimal to maximal congestion in our simulations), AV penetration (AVP) to be within [0, 40] percent (for a realistic amount of controllable AVs in the coming years). The placement of the AVs can either be random or even. For *even placement*, AV are placed every N human-driven vehicles in a lane. For *random placement*, AVs are placed randomly among simulated human-driven vehicles. Merge road geometries include one or two merges at distances that vary between [200, 800] meters, and the main road can have one or two lanes.

3 Learning a robust policy in the single-lane merge scenario

While real-world congestion-reducing driving policies need to operate effectively in a wide variety of traffic conditions, most past research has tested learned policies under the same conditions on which they were trained. Since in the real world it is impractical to deploy a separate policy for each combination of conditions, our primary goal is to understand whether it is feasible to learn a *single* driving policy that is robust to real-world variations in traffic conditions.

The performance of an RL-based driving policy depends on the traffic conditions under which it is trained. We hypothesize that the policy trained under high inflow, medium AV penetration, and random vehicle placement is robust in a range of traffic conditions defined in Section 2 for a single-lane merge scenario. To verify our hypothesis, we discretize the training traffic conditions along their defining dimensions to a total of 30 representative combinations of conditions, as follows. We consider main inflows of 1650, 1850, and 2000 $\mathrm{veh/hour}$ which result in low, medium, and high congestion. We discretize AV placement in traffic to be random or even-spaced. Finally, we discretize the training AV penetration into 5 levels: $10\,\%$, $30\,\%$, $50\,\%$, $80\,\%$, $100\,\%$. Based on this $3 \times 2 \times 5$ discretization, we train 30 policies, one for each combination.

By comparing 30 policies, we verfied our hypothesis and identified a policy that generalizes well across training conditions (which will be termed as identified robust policy). Next, we evaluate the identified robust policy on road geometries different from its training scenario.

4 Deployment to roads with two merging ramps

We first deploy the selected policy on more complex merge roads (with 1500 meters' main road and 250 meters' merge road), which have two merging roads at varying distances (200, 400, 600, or 800 meters), and evaluate the performance of the learned policy with respect to the distance between these two ramps. An example road with two merging on-ramps is shown in Figure 2. We tested the identified robust policy with random AV placement, main inflow of $1800\,\mathrm{veh/hour}$, merge inflow $200\,\mathrm{veh/hour}$, across a range of AV penetrations and the above gaps between the two merging roads. The identified robust policy is better than the human baseline even when the merging ramps are just 200 meters away.

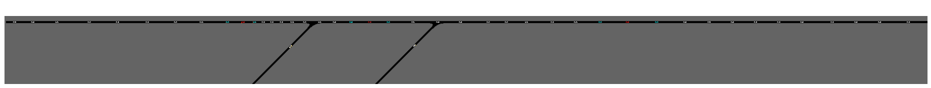

Fig. 2: A more complex road with two merging on-ramps.

5 Deployment to double-lane merge roads

Next, we deploy our identified robust policy on a double-lane merge road by adding a second lane in the main road. Similar to that of the single-lane merge scenario, the vehicles in the right lane must yield to the vehicles from the merging lane and may cause potential congestion in the right lane, while the vehicles in the left lane have the right of way when passing the junction. As a consequence, the vehicles in the left lane tend to move at a faster speed, and there will be more vehicles changing from right to left for speed gain than the number of vehicles changing from left to right. Those lane-changing vehicles cause additional stop-and-go waves in the left lane. To test the robustness of the selected policy in this new road structure, we deploy the learned policy to control the AVs on the right lane. During evaluation, there are only human-driven vehicles in the left lane with inflow $1600\,\mathrm{veh/hour}$, and $10\,\%$ of the vehicles in the right lane are AVs, each of which is controlled by our learned policy. The experimental results shows that the performance of the deployed policy is always significantly better than that of the human-only traffic, regardless of the right main inflow. Hence, the policy trained on the single-lane merge road generalizes well in the double-lane merge scenario.

6 Conclusion

We presented an approach for learning a congestion reduction driving policy that performs robustly in road merge scenarios over a variety of traffic conditions of practical interest. Specifically, the resulting policy reduces congestion in AV penetrations of $1\,\%$–$40\,\%$, traffic inflows ranging from no congestion to heavy congestion, random AV placement in traffic, single-lane single-merge road, single-lane road with two merges at varying distances, and double-lane single-merge road with lane changes. The process of finding this policy involved identifying a single combination of training conditions that yields a robust policy across different evaluating conditions in a single-lane merge scenario. We find, for the first time, that the resulting policy generalizes beyond the training conditions and road geometry it was trained on.

References

1. Cui, J., Macke, W., Yedidsion, H., Goyal, A., Urieli, D., Stone, P.: Scalable multiagent driving policies for reducing traffic congestion. AAMAS (2021)
2. Cummins, L., Sun, Y., Reynolds, M.: Simulating the effectiveness of wave dissipation by followerstopper autonomous vehicles. Transportation Research Part C: Emerging Technologies **123**, 102954 (2021)
3. Lomax, T., Schrank, D., Eisele, B.: 2021 urban mobility report. https://mobility.tamu.edu/umr/, accessed: 2021-10-07
4. Mitarai, N., Nakanishi, H.: Convective instability and structure formation in traffic flow. Journal of the Physical Society of Japan **69**(11), 3752–3761 (2000)
5. Stern, R.E., Cui, S., Delle Monache, M.L., Bhadani, R., Bunting, M., Churchill, M., Hamilton, N., Pohlmann, H., Wu, F., Piccoli, B., et al.: Dissipation of stop-and-go waves via control of autonomous vehicles: Field experiments. Transportation Research Part C: Emerging Technologies **89**, 205–221 (2018)
6. Sugiyama, Y., Fukui, M., Kikuchi, M., Hasebe, K., Nakayama, A., Nishinari, K., ichi Tadaki, S., Yukawa, S.: Traffic jams without bottlenecks—experimental evidence for the physical mechanism of the formation of a jam. New Journal of Physics **10**(3), 033001 (2008)
7. Zhang, Y., Macke, W., Cui, J., Urieli, D., Stone, P.: Learning a robust multiagent driving policy for traffic congestion reduction. arXiv preprint arXiv:2112.03759 (2021)