Toward Socially Friendly Autonomous Driving Using Multi-agent Deep Reinforcement Learning

Extended Abstract

Jhih-Ching Yeh
National Tsing Hua University
Hsinchu, Taiwan
sunny.yeh@gapp.nthu.edu.tw

Von-Wun Soo
Chang Gung University
Taoyuan, Taiwan
soo@cgu.edu.tw

ABSTRACT

We develop a novel multi-agent driving simulation framework (SFDPO) so that socially friendly driving behaviors can be acquired by agents through multi-agent reinforcement learning. We model personal and social driving behaviors in the driver model to reflect human driving goals and preferences. We make a game-theoretic assumption on fair compromised solution concepts to find an equilibrium solution under conflicts in complex interactive scenarios. A meta-policy optimization method is adopted to leverage personal and social driving behaviors in terms of personalized loss and socialized loss to achieve a balanced Pareto optimal solution between the socially friendly and personal preference driving goals.

KEYWORDS

Autonomous Personal and Social Driving Behaviors; MARL; Meta-policy Optimization; Compromised Solution

ACM Reference Format:
Jhih-Ching Yeh and Von-Wun Soo. 2024. Toward Socially Friendly Autonomous Driving Using Multi-agent Deep Reinforcement Learning: Extended Abstract. In *Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), Auckland, New Zealand, May 6 – 10, 2024*, IFAAMAS, 3 pages.

1 INTRODUCTION

As the automobile industry technology advances, the United States Department of Transportation has posted six levels of driving automation defined by the Society of Automotive Engineers (SAE) [10]. It implies that moving vehicles on roads can comprise both human-driving vehicles (HVs) and autonomous vehicles (AVs) without human drivers in the near future as pointed out earlier [4, 5]. HVs encompass diverse characteristics that can classified into personal driving behaviors and social driving behaviors. Personal driving behaviors [9, 18, 22, 24] are actions typically motivated by personal goals and preferences. Social driving behaviors [14, 17, 19, 23] refer to the interactions among AVs or HVs. When some nearby vehicle changes lanes by overtaking, a human driver must respond by an action such as either moving forward, yielding, turning, or stopping. Since AVs without a human driver cannot yet make subtle social driving behaviors coherent with that of HVs, they may cause negative impacts on HVs. It is a challenge for AVs to co-exist with HVs in sharing potentially conflicting road resources [1, 25]. Therefore, autonomous driving is a complex but critical multi-agent decision-making task [2, 13, 16, 20, 21] that deserves investigation.

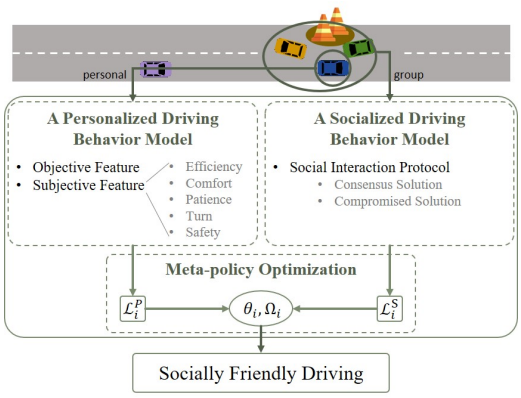

Figure 1: The proposed SFDPO framework.

In general, autonomous driving involves three significant challenges: (1) Personality problem: The difficulty in simulating a variety of personal driving behaviors according to different drivers' goals and preferences that may vary over time. (2) Interactivity problem: The lack of mutual negotiation (compromise) and communication between drivers in order to avoid deadlock or an infinite loop due to potential conflict on using the road resources. (3) Equilibrium problem: A traffic environment with multiple driver encounters may face the problem of finding a compromised solution to balance between personal and social driving behaviors that must be attributed to the driving policies of the drivers. Overall, we argue that AV that can demonstrate well balance between personalized and socially driving behaviors can also lead to a safer, more efficient, as well as more friendly traffic environment. Therefore, it is desirable to develop an efficient and effective algorithm to find the optimal driving policy for the personal and social driving behaviors as a whole to cope with various scenarios.

2 APPROACHES

2.1 Overview of the Proposed Framework

Figure 1 shows the overview of our framework called Socially Friendly Driving Policy Optimization (SFDPO) in which we implement both personalized and socialized driving behavior models, to be introduced in section 2.2 and section 2.3 respectively. The

Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), N. Alechina, V. Dignum, M. Dastani, J.S. Sichman (eds.), May 6 – 10, 2024, Auckland, New Zealand. © 2024 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

personalized driving behavior model discerns personal policies aligned with different goals and preferences, while the socialized behavior model derives social policies in facilitating cooperation or negotiation among others. Moreover, we introduce the meta-policy optimization notion to resolve the equilibrium problem that can balance personal and social policies, to be described in section 2.4.

2.2 Personalized Driving Behavior Model

To simulate human personal driving behaviors, we envision that each agent has a personal preference for certain driving features that may change over time in response to the dynamic environment. Therefore, we conceive the personalized driving behavior model with both objective [6, 9, 12, 14] and subjective driving features [9]. We assume a linear-structured personalized reward function $r^P_{i,t}$ for agent i at time step t that is a weighted sum of selected features as represented in Equation (1). Here, $\boldsymbol{\omega}_{i,t} = [\omega^O_{1,i,t}, \omega^O_{2,i,t}, ..., \omega^O_{M,i,t}, \omega^S_{1,i,t}, \omega^S_{2,i,t}, ..., \omega^S_{N,i,t}]$ is the preference weight vector with M-dimensional objective and N-dimensional subjective weights, while $\mathbf{B}(o_{i,t}, u_{i,t}) = [B^O_1(o_{i,t}, u_{i,t}), B^O_2(o_{i,t}, u_{i,t}), ..., B^O_M(o_{i,t}, u_{i,t}), B^S_1(o_{i,t}, u_{i,t}), B^S_2(o_{i,t}, u_{i,t}), ..., B^S_N(o_{i,t}, u_{i,t})]$ consists of the extracted objective and subjective driving feature function vectors, mapping from specific observations $o_{i,t}$, and actions $u_{i,t}$ to actual rewards.

$$r^P_{i,t} = \mathcal{R}^P(o_{i,t}, u_{i,t}, \boldsymbol{\omega}_{i,t}) = (\boldsymbol{\omega}_{i,t})^T \mathbf{B}(o_{i,t}, u_{i,t}) \tag{1}$$

Based on IPPO [3], we define the personalized loss as shown in Equation (2).

$$\mathcal{L}^P_i(\theta_i, \Omega_i) = -\mathbb{E}_{o,u}\left[\min\left(\rho A^P_{\Omega_i,i,t}, \text{clip}(\rho, 1-\epsilon, 1+\epsilon) A^P_{\Omega_i,i,t}\right)\right] \tag{2}$$

Here, θ_i and Ω_i denote the parameters associated with the policy and generated preference weights models. $A^P_{\Omega_i,i,t}$ is the personalized advantage. The clipped importance sampling factor is defined as $\rho = \frac{\pi_{i,new}(o_{i,t}|u_{i,t})}{\pi_{i,old}(o_{i,t}|u_{i,t})}$, where $\pi_{i,old}$ and $\pi_{i,new}$ are the policies that generate the samples and the updated policy respectively.

2.3 Socialized Driving Behavior Model

The main challenge of determining proper personalized driving behaviors under socially friendly interactions is the acquisition of the weights of subjective driving features. We therefore propose the group-based socialized driving behavior model based on game theory and Nash equilibrium [11]. It aims at acquiring the optimal weights and driving strategy (the group-based joint action) to facilitate friendly interactions among agents. We focus on finding a consensus solution for a group as a Nash equilibrium solution. However, it may not always exist in real scenarios. Thus, we assume a loss-sharing concept to find a compromised solution, where all agents in the same group can come up with a driving strategy combination as the hypothesized consensus solution of all group agents under situations. The loss-sharing compromised solution is to balance agents' expected personalized rewards and losses. It turns out that we can obtain the fairer group-based socialized reward through simulation. It shows that group agents can acquire their optimal driving policies in terms of preference weight vector $\boldsymbol{\omega}_{i,t}$ that best aligns with the personality of agent i and the corresponding scenario while at the same time enhance the group reward. The concept of preference advantage, denoted as $A^{prefer}_{\Omega_i,i,t}$, regarding mapping the subjective driving features into the socialized loss to update Ω_i proposed. The resulting socialized loss is denoted in Equation (3), where $A^S_{i,t}$ is the socialized advantage, and λ_{perfer} is the hyperparameter that can be adjusted to aid optimization.

$$\begin{aligned}\mathcal{L}^S_i(\theta_i, \Omega_i) = \; & - \; \mathbb{E}_{o,\boldsymbol{u}}\left[\min\left(\rho A^S_{i,t}, \text{clip}(\rho, 1-\epsilon, 1+\epsilon) A^S_{i,t}\right)\right] \\ & - \; \lambda_{perfer}\mathbb{E}_{o,\boldsymbol{u}}[\min\left(A^{prefer}_{\Omega_i,i,t}\right)]\end{aligned} \tag{3}$$

Table 1: Comparison against baseline methods.

Scenario	Metrics	IPPO	CoPO	SFDPO
Crossing	Success Rate (%)	59.38	46.88	**100.00**
	Disparity Ratio	1.92	1.51	**1.00**
	Deadlock Number	6.00	23.25	**0.00**

2.4 Meta-policy Optimization

To establish a safe, efficient, and friendly environment, the goal is to determine the optimal driving policy π^*_i for each agent i, corresponding to both parameters θ^*_i and Ω^*_i. This policy is intended not only to minimize both personalized and socialized losses but also to strike a balance between these losses. Thus, we introduce a meta-policy optimization methodology comprising two levels: the object and the meta levels. For the object level, we employ a sequential procedure to update $\mathcal{L}^P(\theta_i, \Omega_i)$ and $\mathcal{L}^S(\theta_i, \Omega_i)$. For the meta level, there are two learners: the base learner, which emphasizes specific losses parameterized by θ_i, and the meta-learner, which ensures that the base learner can adapt to different losses parameterized by Ω_i. Consequently, the meta-policy optimization approach iteratively searches for a balanced Pareto optimal solution.

3 EXPERIMENTS

We conduct the experiments in the crossing scenario [1, 7, 8] and compare the performance of SFDPO against two previous MARL baseline models based on Proximal Policy Optimization (PPO) [15], IPPO [3] and CoPO [14]. Three general-purpose metrics are defined. The success rate is measured as the ratio of the number of agents successfully reaching their goals against the total number of agents participating. The disparity ratio is defined as the ratio of the highest velocity against the lowest one among agents in an episode. Deadlock number is the number of deadlock occurrences in a scenario. As shown Table 1, the proposed SFDPO achieves an average performance improvement against CoPO [14] by an increase of success rate 53.12%, and likewise by a reduction of disparity ratio 33.77% and number of deadlocks 100%.

4 CONCLUSION

SFDPO allows AVs with MARL to acquire proper driving policies that lead to better social driving behaviors in facing with various conflicting scenarios. The meta-policy optimization method tends to leverage personal driving behaviors and social driving behaviors and can find a balanced Pareto optimal solution between two optimization objectives, personalized loss and socialized loss.

REFERENCES

[1] Shunsuke Aoki and Ragunathan (Raj) Rajkumar. 2018. Dynamic Intersections and Self-Driving Vehicles. In *Proceedings of the 9th ACM/IEEE International Conference on Cyber-Physical Systems (ICCPS '18)*. IEEE Press, Porto, Portugal, 320–330. https://doi.org/10.1109/ICCPS.2018.00038

[2] Dian Chen and Philipp Krähenbühl. 2022. Learning from All Vehicles. arXiv:2203.11934 [cs.RO]

[3] Christian Schroeder de Witt, Tarun Gupta, Denys Makoviichuk, Viktor Makoviychuk, Philip H. S. Torr, Mingfei Sun, and Shimon Whiteson. 2020. Is Independent Learning All You Need in the StarCraft Multi-Agent Challenge? arXiv:2011.09533 [cs.AI]

[4] Xuan Di and Rongye Shi. 2020. A Survey on Autonomous Vehicle Control in the Era of Mixed-Autonomy: From Physics-Based to AI-Guided Driving Policy Learning. arXiv:2007.05156 [cs.AI]

[5] Kurt Dresner and Peter Stone. 2007. Sharing the Road: Autonomous Vehicles Meet Human Drivers. In *Proceedings of the 20th International Joint Conference on Artifical Intelligence* (Hyderabad, India) *(IJCAI'07)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1263–1268.

[6] Yali Du, Lei Han, Meng Fang, Ji Liu, Tianhong Dai, and Dacheng Tao. 2019. LIIR: Learning Individual Intrinsic Reward in Multi-Agent Reinforcement Learning. In *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Eds.), Vol. 32. Curran Associates, Inc., Vancouver, Canada. https://proceedings.neurips.cc/paper_files/paper/2019/file/07a9d3fed4c5ea6b17e80258dee231fa-Paper.pdf

[7] Freepik Flaticon. [n.d.]. Top down car icons. https://www.flaticon.com/free-icons/top-down-car

[8] Freepik. [n.d.]. Shop icon. https://www.freepik.com/icon/shop_2981297

[9] Zhiyu Huang, Jingda Wu, and Chen Lv. 2021. Driving Behavior Modeling using Naturalistic Human Driving Data with Inverse Reinforcement Learning. arXiv:2010.03118 [cs.RO]

[10] SAE International. 2021. SAE Levels of Driving Automation™ Refined for Clarity and International Audience. https://www.sae.org/blog/sae-j3016-update

[11] Nail Kashaev and Bruno Salcedo. 2019. Discerning Solution Concepts. arXiv:1909.09320 [econ.EM]

[12] W. Bradley Knox, Alessandro Allievi, Holger Banzhaf, Felix Schmitt, and Peter Stone. 2022. Reward (Mis)design for Autonomous Driving. arXiv:2104.13906 [cs.LG]

[13] Qi Liu, Xueyuan Li, Shihua Yuan, and Zirui Li. 2021. Decision-Making Technology for Autonomous Vehicles Learning-Based Methods, Applications and Future Outlook. arXiv:2107.01110 [cs.RO]

[14] Zhenghao Peng, Quanyi Li, Ka Ming Hui, Chunxiao Liu, and Bolei Zhou. 2022. Learning to Simulate Self-Driven Particles System with Coordinated Policy Optimization. arXiv:2110.13827 [cs.LG]

[15] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal Policy Optimization Algorithms. arXiv:1707.06347 [cs.LG]

[16] Wilko Schwarting, Javier Alonso-Mora, and Daniela Rus. 2018. Planning and Decision-Making for Autonomous Vehicles. *Annu. Rev. Control. Robotics Auton. Syst.* 1 (2018), 187–210. https://api.semanticscholar.org/CorpusID:64853900

[17] Wilko Schwarting, Alyssa Pierson, Javier Alonso-Mora, Sertac Karaman, and Daniela Rus. 2019. Social behavior for autonomous vehicles. *Proceedings of the National Academy of Sciences of the United States of America* 116 (2019), 24972 – 24978. https://doi.org/10.1073/pnas.1820676116

[18] David Shinar and Ilit Oppenheim. 2011. Review of Models of Driver Behaviour and Development of a Unified Driver Behaviour Model for Driving in Safety Critical Situations. In *Human Modelling in Assisted Transportation*, P. Carlo Cacciabue, Magnus Hjälmdahl, Andreas Luedtke, and Costanza Riccioli (Eds.). Springer Milan, Milano, 215–223. https://doi.org/10.1007/978-88-470-1821-1_23

[19] Von-Wun Soo. 2000. Agent Negotiation in Trusted Third Party Mediated Uncertain Games. In *Proceedings of the Fourth International Conference on Autonomous Agents* (Barcelona, Spain) *(AGENTS '00)*. Association for Computing Machinery, New York, NY, USA, 265–266. https://doi.org/10.1145/336595.337482

[20] Ming Tan. 1997. *Multi-Agent Reinforcement Learning: Independent vs. Cooperative Agents*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 487–494.

[21] Dimitrios Troullinos, Georgios Chalkiadakis, Ioannis Papamichail, and Markos Papageorgiou. 2021. Collaborative Multiagent Decision Making for Lane-Free Autonomous Driving. In *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems* (Virtual Event, United Kingdom) *(AAMAS '21)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1335–1343.

[22] Junhua Wang, Wenxiang Xu, Ting Fu, Hongren Gong, Qiangqiang Shangguan, and Anae Sobhani. 2022. Modeling aggressive driving behavior based on graph construction. *Transportation Research Part C: Emerging Technologies* 138 (2022), 103654. https://doi.org/10.1016/j.trc.2022.103654

[23] Wenshuo Wang, Letian Wang, Chengyuan Zhang, Changliu Liu, and Lijun Sun. 2022. Social Interactions for Autonomous Driving: A Review and Perspective. *Found. Trends Robotics* 10 (2022), 198–376. https://doi.org/10.1561/2300000078

[24] Yunpeng Wang, Junjie Zhang, and Guangquan Lu. 2019. Influence of Driving Behaviors on the Stability in Car Following. *IEEE Transactions on Intelligent Transportation Systems* 20 (2019), 1081–1098. https://doi.org/10.1109/TITS.2018.2837740

[25] Moritz Werling, Julius Ziegler, Sören Kammel, and Sebastian Thrun. 2010. Optimal trajectory generation for dynamic street scenarios in a Frenét Frame. *2010 IEEE International Conference on Robotics and Automation* (2010), 987–993. https://api.semanticscholar.org/CorpusID:9467550