Communication and Generalization in Multi-Agent Learning

Doctoral Consortium

Jiaxun Cui
The University of Texas at Austin
Austin, TX, United States
cuijiaxun@utexas.edu

ABSTRACT

Significant challenges exist in robustly interacting and communicating with a diverse array of agents, especially in intricate settings like autonomous driving where AI agents and humans coexist. This work approaches these challenges from three perspectives: generalization of agent policies, development of communication-supporting representations, and interactions between humans and AI agents using natural language. We provide an overview of preliminary achievements in each area and outline proposed research focusing on enhancing cooperative driving through natural language communication, aiming to comprehensively address these complex multi-agent interaction challenges.

KEYWORDS

Communication; Generalization; Human-AI Interaction; LLMs

ACM Reference Format:

Jiaxun Cui. 2024. Communication and Generalization in Multi-Agent Learning: Doctoral Consortium. 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

Future AI is not alone. Artificial Intelligence (AI) has been intensively developed in the single-agent setting such as playing Atari games and generating artistic pieces. It is relatively under-explored in scenarios where decision-making AI agents coexist with humans and other agents. For instance, there is a potential for autonomous vehicles to be able to work together to improve traffic safety and cooperate with human drivers to reduce traffic congestion [11].

The learning dynamics in multi-agent decision-making scenarios present a significant challenge. When the other agents are viewed as a part of the environment, the environment is non-stationary. Game theory offers some analytical tools to solve multi-agent games, but struggles with high game complexity and handling multi-modal (in formats like videos, sounds, HDMaps, Metadata, etc.) input for policies. Multi-agent Reinforcement Learning (MARL), which optimizes expected individual or team return using Reinforcement Learning (RL), provides promising scalability for complex inputs and long-horizon decisions. Techniques in MARL like population-based training, empirical game-theoretic analysis, and centralized training decentralized execution (CTDE), have achieved significant success in games like Go [9], StarCraft [12], and Diplomacy [1].

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.

Despite MARL's ability to formulate strong policies that yield high expected returns, two specific aspects are yet to be extensively examined in the context of deep learning.

The first aspect is the generalization of a policy to emergent teammates or opponents that do not appear during the training phase, herein referred to as *unseen* agents. The cooperative counterpart of the problem is referred to as Ad-Hoc Teamwork (AHT) [10] or Zero-Shot Coordination (ZSC) [6]. This concern is crucial since AI may frequently encounter novel partners, such as humans or other AI agents. These partners might exhibit diverse policy styles or coordination conventions. Typically, the generalization of a policy is attained by either calculating conservative policies such as Nash Equilibria or by training with a diverse population of agents.

The second aspect involves communication in multi-agent learning. Most training frameworks assume either entirely centralized information sharing or fully decentralized observation. However, real-world scenarios often permit communication, albeit with limitations in bandwidth. These restrictions complicate multi-agent learning, raising critical questions about when and whom to communicate with, what information to share, and how to handle the received information. Among communication protocols, when interacting with humans, natural language is the most suitable. It is well-structured and can be understood by almost all humans, perhaps with proper translation.

Where these two dimensions intersect, formulating a universally applicable policy that accommodates a range of policies, along with communication mechanisms, holds practical relevance in real-life applications such as autonomous driving. Consider a scenario where a vehicle, due to a brake failure, is about to run a red light and broadcasts its intention to nearby vehicles as a cautionary measure. A vehicle intending to cross at the green light could receive the message and decelerate accordingly. Note that if one of the cars is controlled by a human driver, the message may need to be in natural language, or some other human-interpretable modality.

With this motivation in mind, this work navigates the complexities of multi-agent learning, addressing the critical question:

How can a decision-making agent efficiently communicate with and create generalizable policies for novel AI or human teammates or opponents in simulated real-world scenarios?

This work will explore and answer the question along the following three dimensions:

(1) **Generalization**. Such generalization pertains to interacting with diverse partners or opponents without the need for additional fine-tuning during interactions. Ensuring a high degree of policy population *diversity* during the training phase prepares the agents to handle a broad spectrum of policy styles and coordination conventions.

(2) **Communication-Supporting Representations**. Interchange of messages involves answering questions related to when, with whom, and what to communicate, and managing the tangible real-world constraints of limited bandwidth. This work will focus on how to construct messages through a learned representation space in decision-making scenarios.
(3) **Interaction with Humans and other AI agents via Natural Language.** This work will deepen the investigation into the intersection of communication and policy generalization with humans and other AI agents. To this end, we plan to integrate natural language into the multi-agent system so that agents can share their intentions and key observations with both human and autonomous agents. Concurrently, humans will have the opportunity to strategize with nearby agents to optimize specific metrics.

All proposed contributions will be evaluated in simulated environments that model real-world applications following the best practices in the literature.

2 PRELIMINARY RESULTS

In this section, we outline some preliminary results along each dimension of the main question:

We introduce and evaluate a reinforcement and game-theoretic training framework, MACTA [5], which uses PPO [8] as the best response oracle and fictitious play [2] as the empirical game-theoretic tool. We show that the resultant policy can generalize to unseen opponents and is robust against a dedicated adaptive opponent in a simulated cache timing attack scenario. Further, we propose the concept of *Minimum Coverage Sets* (MCS) in [7] as a metric for maximizing the meaningful diversity in the training population. We also propose the algorithm L-BRDiv to approximate the MCS in a game. Results show that we can robustly discover distinct cooperative conventions and Ad Hoc Teamwork agents trained with the generated set of diverse populations can adapt robustly to a variety of testing agents. These two works partially address the generalization dimension (Dimension 1).

We introduce and evaluate Coopernaut [4], an end-to-end driving framework that generates efficient communicatable representations of the local point-cloud observation of autonomous agents through end-to-end imitation learning of driving policies. We show that with Coopernaut, autonomous agents can significantly reduce collisions without compromising traffic efficiency compared to disconnected vehicles in accident-prone scenarios. This contribution addresses *what* information to communicate (Dimensions 2) through a learned representation space under the available bandwidth in autonomous driving.

We have conducted an empirical study [3] of applying decentralized multi-agent reinforcement learning to work with both humans and AI agents to improve traffic efficiency in autonomous driving. We delve into the decentralized training of RL agents in a mixed environment where human and AI agents coexist. Experimental findings indicate that a small presence of RL autonomous vehicles can effectively collaborate to influence human drivers and amplify overall traffic efficiency within an open environment. This contribution explores the agents' intelligent interactions with both humans and other AI agents (Dimension 3) without communication.

3 PROPOSED RESEARCH

In the evolving landscape of traffic systems, autonomous vehicles (AVs) must interact not only with human drivers but also with other AVs from various manufacturers, each potentially adhering to different driving policies. Thus, autonomous vehicles being able to communicate and form emergent coordination with novel partners through a generalizable communication language (e.g. natural language) is one promising future direction in traffic systems.

Natural language serves as a well-structured and universally comprehensible communication protocol among humans, adaptable with appropriate translation. By equipping AVs with a language module, they can communicate their intentions or observations to humans, as well as respond to instructions received in natural language. For instance, an AV experiencing brake failure and consequently running a red light can broadcast its predicament to nearby vehicles. A vehicle preparing to cross the intersection on a green light, upon receiving this message, can then slow down to avoid a potential collision. Different from the generated representations in Coopernaut, the natural languages are more interpretable. Furthermore, the integration of large language models (LLMs) enhances the capabilities of these autonomous agents. LLMs enable AVs to perform common-sense analysis based on messages received and engage in meaningful dialogues with multiple vehicles, thus enriching the context and decision-making process in driving.

Addressing policy diversity in autonomous driving is another critical aspect of this project. Different driving styles, such as risk-seeking or risk-averse policies, significantly impact decision-making in scenarios like overtaking. A risk-seeking AV might prioritize reducing travel time at the risk of increased collision potential, whereas a risk-averse AV might opt for safety but risk getting stuck in a deadlock. We plan to have vehicles negotiate about their plans to reach local Ad Hoc coordination for improved traffic efficiency.

This proposed project is dedicated to achieving significant advancements in the domain of autonomous vehicle communication and coordination. Our goals, leveraging the utility of natural language, are multifaceted and aim to enhance the interaction dynamics within diverse traffic scenarios. These goals include:

(1) Let AVs articulate their intentions and critical observations to other vehicles effectively. In scenarios where conflicts may arise, we aim to enable vehicles to negotiate and collaboratively formulate driving plans to avoid conflicts.
(2) Integrate LLMs in AVs to interpret received messages and pivotal driving-related information to generate language instructions for driving decisions for AVs to follow.
(3) Equip agents with the ability to interact seamlessly with a range of driving styles and policies, including behaviors of human drivers, ensuring that AVs can operate safely and efficiently in a mixed-autonomy traffic environment.
(4) Refine driving policies of AVs to achieve a balance of high driving efficiency, a low rate of collisions, and rapid adaptability to varying driving policies.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to Prof. Peter Stone for his invaluable supervision, guidance, and support throughout this thesis research.

REFERENCES

[1] Anton Bakhtin, Noam Brown, Emily Dinan, Gabriele Farina, Colin Flaherty, Daniel Fried, Andrew Goff, Jonathan Gray, Hengyuan Hu, et al. 2022. Human-level play in the game of Diplomacy by combining language models with strategic reasoning. *Science* 378, 6624 (2022), 1067–1074.

[2] George W Brown. 1951. Iterative solution of games by fictitious play. *Act. Anal. Prod Allocation* 13, 1 (1951), 374.

[3] Jiaxun Cui, William Macke, Harel Yedidsion, Daniel Urieli, and Peter Stone. 2021. Scalable multiagent driving policies for reducing traffic congestion. *arXiv preprint arXiv:2103.00058* (2021).

[4] Jiaxun Cui, Hang Qiu, Dian Chen, Peter Stone, and Yuke Zhu. 2022. Coopernaut: end-to-end driving with cooperative perception for networked vehicles. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 17252–17262.

[5] Jiaxun Cui, Xiaomeng Yang, Mulong Luo, Geunbae Lee, Peter Stone, Hsien-Hsin S. Lee, Benjamin Lee, G. Edward Suh, Wenjie Xiong, and Yuandong Tian. 2023. MACTA: A Multi-agent Reinforcement Learning Approach for Cache Timing Attacks and Detection. In *The Eleventh International Conference on Learning Representations*. https://openreview.net/forum?id=CDlHZ78-Xzi

[6] Hengyuan Hu, Adam Lerer, Alex Peysakhovich, and Jakob Foerster. 2020. "other-play" for zero-shot coordination. In *International Conference on Machine Learning*. PMLR, 4399–4410.

[7] Arrasy Rahman, Jiaxun Cui, and Peter Stone. 2023. Minimum coverage sets for training robust ad hoc teamwork agents. *arXiv preprint arXiv:2308.09595* (2023).

[8] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal Policy Optimization Algorithms. *CoRR* abs/1707.06347 (2017). arXiv:1707.06347 http://arxiv.org/abs/1707.06347

[9] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, et al. 2017. Mastering the game of go without human knowledge. *nature* 550, 7676 (2017), 354–359.

[10] Peter Stone, Gal Kaminka, Sarit Kraus, and Jeffrey Rosenschein. 2010. Ad hoc autonomous agent teams: Collaboration without pre-coordination. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 24. 1504–1509.

[11] Eugene Vinitsky, Aboudy Kreidieh, Luc Le Flem, Nishant Kheterpal, Kathy Jang, Cathy Wu, Fangyu Wu, Richard Liaw, Eric Liang, and Alexandre M Bayen. 2018. Benchmarks for reinforcement learning in mixed-autonomy traffic. In *Conference on Robot Learning*. 399–409.

[12] Oriol Vinyals, Igor Babuschkin, Wojciech M Czarnecki, Michaël Mathieu, Andrew Dudzik, Junyoung Chung, David H Choi, Richard Powell, Timo Ewalds, Petko Georgiev, et al. 2019. Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature* 575, 7782 (2019), 350–354.