Explanations for Multi-Agent Reinforcement Learning

Kayla Boggess

School of Engineering and Applied Science, University of Virginia
kjb5we@virginia.edu

Abstract

Explainable reinforcement learning (xRL) provides explanations for "black-box" decision making systems. However, most work in xRL is based on single-agent settings instead of the more complex multi-agent reinforcement learning (MARL). Several different types of post-hoc explanations must be provided to increase understanding of both *centralized* and *decentralized* MARL systems. For centralized MARL, this research develops methods to generate *global policy summaries*, *query-based explanations*, and *temporal explanations*. For decentralized MARL, this research develops *global policy summaries* and *query-based explanations*.

Introduction

Explainable reinforcement learning (xRL) provides explanations for the behaviors of "black-box" decision making systems. However, most work in xRL is based on single-agent settings instead of the more complex multi-agent reinforcement learning (MARL) (Heuillet, Couthouis, and Díaz-Rodríguez 2021). By providing post-hoc (after policy training) explanations regarding more complex multi-agent actions, we can facilitate human-agent collaboration, increasing system transparency and user satisfaction (Kraus et al. 2020) for exciting new applications and domains (e.g., search and rescue). Overall, this research considers two main objectives: *generate explanations for centralized MARL and generate explanations for decentralized MARL*. For centralized MARL, this research develops methods to generate *global policy summaries*, *query-based explanations*, and *temporal explanations*. For decentralized MARL, this research develops *global policy summaries* and *query-based explanations*.

Overview and Expected Contributions

Several different types of explanations must be provided to increase understanding of MARL systems. First, *policy summarizations* can explain global agent behaviors. Furthermore, *query-based explanations* ("Why don't agents do specific actions in certain states?") and *temporal explanations* ("Why don't agents complete task 1, then task 2, and eventually task 3?") give insight into more local behaviors in specific states. Furthermore, there are two ways MARL policies can be executed, *centralized*, where agents share one joint policy with joint states and actions, and *decentralized*, in which agents have individual policies with individual states and actions (Albrecht, Christianos, and Schäfer 2024). Each suffers from its own challenges regarding post-hoc explanation generation and specific care must be given to develop effective methods for each combination of policy and explanation type.

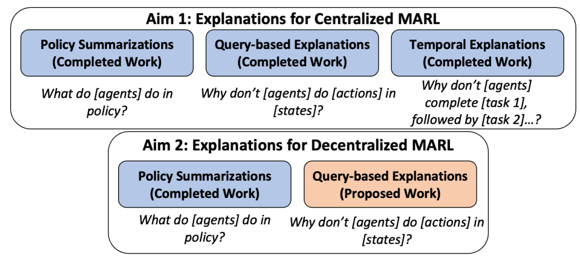

Figure 1: Research Overview

Centralized MARL

Challenges. Generating explanations for centralized MARL comes with several challenges not present in single-agent domains. First, the combinatorial nature of MARL (growing joint state/action space with each additional agent) leads to scalability issues. Furthermore, when producing an explanation, we must provide adequate information about agent behavior, specifically about agent cooperation, but not so much redundant information that it overwhelms the user, leading to a cognitive burden. Finally, presenting the explanation to the user can be difficult when a large number of agents, and thus, a large amount of information is present.

Policy Summarizations and Query-based Explanations. To generate policy summarizations of the most probable sequence of agent behavior and query-based explanations (What?, When?, Why not?), my method begins by building an abstract representation of the policy as a multi-agent Markov decision process (MMDP) via observed policy evaluation samples, adapted from the method presented in (Topin and Veloso 2019) for single-agent RL. The summarization is obtained by isolating the most probable path through the MMDP and presenting the sequence of agent cooperation on tasks. Regarding query-based explanations, my method isolates states from the MMDP satisfying given query criteria, finds a minimal Boolean logic expression to cover them, and converts that logic to a natural language explanation, inspired by the single-agent method in (Hayes and Shah 2017). However, my method uses domain knowledge about possible agent cooperation to counteract scalability issues. These methods were applied to three MARL domains and can effectively produce summarizations and explanations for up to 19 agents. Furthermore, in a study with real-world users, these methods improve user performance and subjective explanation goodness ratings. More information about centralized MARL challenges and the proposed methods can be found in (Boggess, Kraus, and Feng 2022).

Temporal Explanations. For temporal queries, my approach improves on single-agent methods such as (Sreedharan et al. 2022). I generate an abstract policy representation as an MMDP and convert the user's given query into a PCTL* logic formula. The method then uses probabilistic model checking to ensure the query is feasible under the given policy. If the query is infeasible, the method generates a correct and complete explanation as to why. This approach can generate explanations for up to 9 agents in four domains and improves user performance and subjective goodness metrics. More information regarding this method can be found in (Boggess, Kraus, and Feng 2023).

Decentralized MARL

Challenges. Decentralized MARL suffers from the same challenges as centralized MARL but is also burdened with partial visibility (Albrecht, Christianos, and Schäfer 2024). Since no one agent has a total view of the environment, specific interest must be paid to how agent information is processed and aggregated as we do not want to produce only a partial picture of the policy or produce behaviors that do not exist. Furthermore, the decentralization can affect the presentation of explanations since the information from disparate agents can overwhelm users. To our knowledge, this is the first post-hoc approach for generating summarizations and query-based explanations for decentralized MARL.

Policy Summarizations. My method for generating a decentralized policy summarization of the most probable set of agent actions relies on aggregating information from disparate agents via Hasse diagrams (Sarkar 2017), which present agent cooperation in their nodes and the non-determinism of policy execution in their edges. Furthermore, this method is supported by a new presentation technique for decentralized MARL called position overlay, which can be utilized in augmented reality (AR). This method shows effective scalability on multiple MARL domains and increased user performance and satisfaction in an AR-based user study, supported by another student's work.

Query-based Explanations. Currently, I am developing a method to generate query-based explanations (When?, What?, Why not?) for decentralized policies via Hasse diagrams. Additionally, these queries will provide a summary envelope containing rules for all possible agent behaviors under a given policy. Yet, these explanations may contain some redundant information due to their decentralized nature, leading to lengthy explanations and cognitive burden. To remedy this, I would like to use large language models to condense the information using prompt engineering to increase user satisfaction, lower cognitive load, and decrease response time while avoiding altering crucial information due to potential hallucinations. Finally, extensive computational experiments and user studies are needed.

Task	Timeline
Query-based Explanation Method	Jul '24 - Jan '25
Explanation Condensing	Nov '25 - Mar '25
Computational Experiments	Mar '25 - Apr '25
User Studies	Mar '25 - Apr '25

Table 1: Research Timeline

Acknowledgments

This work was supported by U.S. NSF grant CCF-1942836 and U.S. ONR grant N00014-18-1-2829.

References

Albrecht, S. V.; Christianos, F.; and Schäfer, L. 2024. *Multi-Agent Reinforcement Learning: Foundations and Modern Approaches*. MIT Press.

Boggess, K.; Kraus, S.; and Feng, L. 2022. Toward Policy Explanations for Multi-Agent Reinforcement Learning. In *International Joint Conference on Artificial Intelligence (IJCAI)*.

Boggess, K.; Kraus, S.; and Feng, L. 2023. Explainable Multi-Agent Reinforcement Learning for Temporal Queries. In *International Joint Conference on Artificial Intelligence (IJCAI)*.

Hayes, B.; and Shah, J. A. 2017. Improving robot controller transparency through autonomous policy explanation. In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*, 303–312. IEEE.

Heuillet, A.; Couthouis, F.; and Díaz-Rodríguez, N. 2021. Explainability in deep reinforcement learning. *Knowledge-Based Systems*, 214: 106685.

Kraus, S.; Azaria, A.; Fiosina, J.; Greve, M.; Hazon, N.; Kolbe, L.; Lembcke, T.-B.; Muller, J. P.; Schleibaum, S.; and Vollrath, M. 2020. AI for explaining decisions in multi-agent environments. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 13534–13538.

Sarkar, S. K. 2017. *A Textbook of Discrete Mathematics*, chapter 9.4 "Hasse Diagram", 339–341. S. Chand Publishing, 9th edition.

Sreedharan, S.; Soni, U.; Verma, M.; Srivastava, S.; and Kambhampati, S. 2022. Bridging the Gap: Providing Post-Hoc Symbolic Explanations for Sequential Decision-Making Problems with Inscrutable Representations. In *International Conference on Learning Representations*.

Topin, N.; and Veloso, M. 2019. Generation of policy-level explanations for reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 2514–2521.