Towards Explaining Sequences of Actions in Multi-Agent Deep Reinforcement Learning Models

Extended Abstract

Khaing Phyo Wai
Singapore Management University
Singapore
khaingpw@smu.edu.sg

Minghong Geng
Singapore Management University
Singapore
mhgeng.2021@phdcs.smu.edu.sg

Budhitama Subagdja
Singapore Management University
Singapore
budhitamas@smu.edu.sg

Shubham Pateria
Singapore Management University
Singapore
shubhamp@smu.edu.sg

Ah-Hwee Tan
Singapore Management University
Singapore
ahtan@smu.edu.sg

ABSTRACT

Although Multi-agent Deep Reinforcement Learning (MADRL) has shown promising results in solving complex real-world problems, the applicability and reliability of MADRL models are often limited by a lack of understanding of their inner workings for explaining the decisions made. To address this issue, this paper proposes a novel method for explaining MADRL by generalizing the sequences of action events performed by agents into high-level abstract strategies using a spatio-temporal neural network model. Specifically, an interval-based memory retrieval procedure is developed to generalize the encoded sequences of action events over time into short sequential patterns. In addition, two abstraction algorithms are introduced, one for abstracting action events across multiple agents and the other for further abstracting the episodes over time into short sequential patterns, which can then be translated into symbolic form for interpretation. We evaluate the proposed method using the StarCraft Multi Agent Challenge (SMAC) benchmark task, which shows that the method is able to derive high-level explanations of MADRL models at various levels of granularity.

KEYWORDS

Multi Agent Deep Reinforcement Learning; Explainable Artificial Intelligence; Explainable Deep Reinforcement Learning.

ACM Reference Format:
Khaing Phyo Wai, Minghong Geng, Budhitama Subagdja, Shubham Pateria, and Ah-Hwee Tan. 2023. Towards Explaining Sequences of Actions in Multi-Agent Deep Reinforcement Learning Models: Extended Abstract. In *Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), London, United Kingdom, May 29 – June 2, 2023*, IFAAMAS, 3 pages.

1 INTRODUCTION

Multi-agent Deep Reinforcement Learning (MADRL) [5, 6, 22] has been demonstrated to solve complex real-world problems. However, MADRL models use black-box neural networks [17, 25, 29, 31, 39], making the explanation of the learned knowledge contained in such models challenging [9, 14, 23]. Explainable AI (XAI) systems or models aim to offer the predictions in a form that is explainable and understood by humans [1, 3, 21] without a significant amount of additional knowledge from the human participants [15, 16, 33].

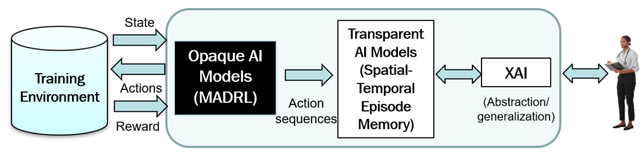

Figure 1: A general framework for explaining opaque MADRL models using transparent self-organizing neural network models and post-hoc abstraction algorithms.

Under the umbrella of XAI, various methods have been proposed for explaining Deep Reinforcement Learning (DRL) models [8, 11, 18, 27, 36]. These methods differ in the form of interpretation and explanation of the implicit knowledge in the DRL models [2, 10, 24]. Some recent work have discussed how low-level sequences of actions can be converted to a high-level explainable form [28, 34]. Although these works are able to abstract agent's action sequences, they are not applicable to the multi-agent domain. On the other hand, Heuillet et al. have proposed using Shapley Values for MADRL model interpretation [20], leading to a general approach for explaining cooperative strategies and agent contribution in a multi-agent system. However, Shapley values cannot explain the actions taken by the agents and the episodic behaviour [7, 12].

This study aims to explain the behavior of MADRL models through interpreting the sequences of action events performed across multiple agents over time. Specifically, we adopt an *explanation by simplification* [19] approach which translates the sequences of low-level primitive action (called episodes) into short and high-level sequences (called abstracted episodes). For empirical evaluation, we have conducted extensive experiments using the SMAC [26, 38, 40] game environment. The results show that the proposed approach can simplify the sequences of action events performed by multiple agents and transform them into comprehensible strategies. More importantly, the abstracted and simplified strategies could potentially be adopted by another multi-agent team to produce performance comparable to those of the original MADRL agents in playing the game.

2 APPROACH AND METHODOLOGY

Our approach to explaining MADRL models is by transferring the learned behaviour into a more transparent episodic memory model

Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), A. Ricci, W. Yeoh, N. Agmon, B. An (eds.), May 29 – June 2, 2023, London, United Kingdom.

and explaining in terms of the sequence of action events performed by the multiple agents. As illustrated in Figure 1, we propose a general framework, wherein the knowledge and behaviour of the black-box MADRL agents (in the forms of the sequences of action events performed by the agents over time) are transferred into a collective sequential memory model based on spatio-temporal neural networks. The memory model, considered as a more transparent AI model, also serves to generalize the sequences of action events across multiple episodes before they are further simplified and abstracted by XAI algorithms for interpretation. Accordingly, the research problems to be addressed can be defined as (1) to design a sequential memory model to learn and abstract the collective trajectories and action events of multi-agent units and (2) to interpret the spatio-temporal action patterns across multiple agents learned by the sequential memory model in symbolic form.

To this end, we present a spatio-temporal neural network model based on a modification of Episodic Memory–Adaptive Resonance Theory (EM-ART) model [32, 37] for encoding and generalizing sequences of action events performed by MADRL agents *across multiple episodes*. In addition, by encoding the time stamp of the action events explicitly using complement coding [30] in a time field [13], an interval-based memory retrieval procedure is developed to generalize the encoded actions and behaviour patterns of the agents over a selected time interval into short abstract sequential patterns. Previously used for retrieving events over a selected time interval in Spatio-Temporal Episodic Memory (STEM) model [4], the *interval-based memory retrieval* method averages the features values of all events occurring within the time interval so as to abstract and discern dominant events. For tuning the degree of abstraction over time, the method incorporates a parameter called *abstraction factor* which can be chosen according to the length of abstracted sequences (and time interval) as preferred by the user.

Furthermore, two new algorithms called Significant Event Selection for Episode Abstraction (SESEA) and Repeated Event Reduction for Episode Abstraction (REREA) are presented for abstracting the episodes of action events encoded by EM-ART into *sequences of significant action events* and *sequences of unique action events* respectively. The combination of various abstraction algorithms serves to transform the long sequences of action events performed by MADRL models into short abstract episodes of notable events (strategies).

3 EXPERIMENTS

Our experiments are conducted using the StarCraft Multi-Agent Challenge (SMAC) [26] which is a set of mini-games developed based on the StarCraft II (SC2) [35, 41] real-time strategy (RTS) game platform. We consider a combat scenario called **4t**, wherein two groups of identical units, namely the allied group trained by MADRL model and the enemy group based on built-in rules, are symmetrically placed on the map and the objective of the game is to maximize the win rate. The agents are homogeneous in the reported experiments wherein the type of actions they can perform are the same. To train the MADRL team, we employ QMIX [25], which has produced competitive performance in the SMAC environment.

In our experiments, a collection of 1,100 episodes performed by MADRL model is recorded after learning is completed. Subsequently, we built EM-ART models using various EM-ART vigilance parameter values to study their effects on learning and generalization of events and episodes. Table 1 shows a set of abstracted events from winning episode extracted by the EM-ART model using the interval-based memory retrieval algorithm with an abstraction factor of 10. Based on the extracted episode, the two abstraction algorithms are further applied and the final episode abstracted over agents and time is shown in Table 2.

Time Interval	Action	Time Interval	Action
t1-t3	WWWW	t16-t18	$A_3A_3A_3A_3$
t4-t6	SNNS	t19-t21	$A_1A_3A_1A_1$
t7-t9	WWWW	t22-t24	$A_1A_1A_2A_1$
t10-t12	EEEE	t25-t27	$A_2A_2A_2A_2$
t13-t15	$A_0A_0A_0A_0$	t28-t30	$A_2A_2A_2$E

Table 1: A set of abstracted events extracted by EM-ART using time-based memory retrieval. Legend of actions: N, S, E, and W indicate move[north], move[south], move[east], and move[west], respectively; A_i indicates attack[enemy_i]; - indicates stop; and X indicates no_op.

Time Interval	Action	Time Interval	Action
t1-t3	W	t13-t15	A_0
t4-t6	SN	t16-t18	A_3
t7-t9	W	t19-t24	A_1
t10-t12	E	t25-t30	A_2

Table 2: A winning strategy derived with event abstraction over agents and episode abstraction over time.

We note that the sequence of low-level actions performed by the MADRL agents can be explained in terms of high-level abstracted strategies as follows. At the group level, the multi-agent team performs a series of the move actions before proceeding with the attack actions. At the subgroup level, agents also perform strategic positioning before engaging in firing. During the combat, the agents perform coordinated attack on the same enemy unit by always attacking enemies in a specific order (A_0, A_3, A_1 followed by A_2). Though the reported evaluation is based on the **4t** scenario, our method can be used to explain both single agent and multi-agent models and can be scaled according to the number of agents.

4 CONCLUSION

To explain how a team of trained MADRL agents selects their actions cooperatively, a spatial-temporal memory called EM-ART can be used to encode the sequences of actions performed by the team of decentralized agents trained using the MADRL model. More importantly, the action sequences can be further generalized and abstracted using the time-based retrieval method and the two episode abstraction algorithms. The results have shown that the proposed approach is able to explain the behavior of the multi-agent team by means of the generalized and abstracted episodes. The developed methods are also generalizable to other scenarios as the abstraction factor (across time) and the level of abstraction (across agents) parameters enable the methods to be applicable to episodes of different length and any number of agents. Beyond explaining the action sequences observed, it will be interesting in the future to also consider contextual information captured in the state attributes, for further explaining the decision-making process of MADRL agents.

ACKNOWLEDGMENTS

This research was supported by the DSO National Laboratories, Singapore (Agreement No. DSOCL20200) and the Jubilee Technology Fellowship awarded to Ah-Hwee Tan by Singapore Management University.

REFERENCES

[1] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina, Richard Benjamins, et al. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information fusion* 58 (2020), 82–115.

[2] Angel Ayala, Francisco Cruz, Bruno J. T. Fernandes, and Richard Dazeley. 2021. Explainable Deep Reinforcement Learning Using Introspection in a Non-episodic Task. *CoRR* abs/2108.08911 (2021). arXiv:2108.08911

[3] Davide Calvaresi, Amro Najjar, Michael Winikoff, and Kary Främling. 2020. *Explainable, Transparent Autonomous Agents and Multi-Agent Systems*. Springer.

[4] Poo-Hee Chang and Ah-Hwee Tan. 2017. Encoding and recall of spatio-temporal episodic memory in real time. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*. 1490–1496.

[5] Zichen Chen, Budhitama Subagdja, and Ah-Hwee Tan. 2019. End-to-end deep reinforcement learning for multi-agent collaborative exploration. In *2019 IEEE International Conference on Agents (ICA)*. IEEE, 99–102.

[6] Wei Du and Shifei Ding. 2021. A survey on multi-agent deep reinforcement learning: from the perspective of challenges and applications. *Artificial Intelligence Review* 54, 5 (2021), 3215–3238.

[7] IU Ekanayake, DPP Meddage, and Upaka Rathnayake. 2022. A novel approach to explain the black-box nature of machine learning in compressive strength predictions of concrete using Shapley additive explanations (SHAP). *Case Studies in Construction Materials* 16 (2022), e01059.

[8] Vilde B Gjærum, Inga Strümke, Jakob Løver, Timothy Miller, and Anastasios M Lekkas. 2023. Model tree methods for explaining deep reinforcement learning agents in real-time robotic applications. *Neurocomputing* 515 (2023), 133–144.

[9] Sven Gronauer and Klaus Diepold. 2022. Multi-agent deep reinforcement learning: a survey. *Artificial Intelligence Review* (2022), 1–49.

[10] Wenbo Guo, Xian Wu, Usmann Khan, and Xinyu Xing. 2021. Edge: Explaining deep reinforcement learning policies. *Advances in Neural Information Processing Systems* 34 (2021), 12222–12236.

[11] Alexandre Heuillet, Fabien Couthouis, and Natalia Díaz-Rodríguez. 2021. Explainability in deep reinforcement learning. *Knowledge-Based Systems* 214 (2021), 106685.

[12] Alexandre Heuillet, Fabien Couthouis, and Natalia Díaz-Rodríguez. 2022. Collective explainable AI: Explaining cooperative strategies and agent contribution in multiagent reinforcement learning with shapley values. *IEEE Computational Intelligence Magazine* 17, 1 (2022), 59–71.

[13] Yue Hu, Budhitama Subagdja, Ah-Hwee Tan, Chai Quek, and Quanjun Yin. 2022. Who are the 'silent spreaders'?: Contact tracing in spatio-temporal memory models. *Neural Computing and Applications* 34, 17 (2022), 14859–14879.

[14] Stefan Kolek, Duc Anh Nguyen, Ron Levie, Joan Bruna, and Gitta Kutyniok. 2022. A rate-distortion framework for explaining black-box model decisions. In *xxAI-Beyond Explainable AI: International Workshop, Held in Conjunction with ICML 2020, July 18, 2020, Vienna, Austria, Revised and Extended Papers*. Springer, 91–115.

[15] Theocharis Kravaris, Konstantinos Lentzos, Georgios Santipantakis, George A Vouros, Gennady Andrienko, Natalia Andrienko, Ian Crook, Jose Manuel Cordero Garcia, and Enrique Iglesias Martinez. 2022. Explaining deep reinforcement learning decisions in complex multiagent settings: towards enabling automation in air traffic flow management. *Applied Intelligence* (2022), 1–36.

[16] Deepthi Praveenlal Kuttichira, Sunil Gupta, Cheng Li, Santu Rana, and Svetha Venkatesh. 2019. Explaining Black-Box Models Using Interpretable Surrogates. In *PRICAI 2019: Trends in Artificial Intelligence: 16th Pacific Rim International Conference on Artificial Intelligence, Cuvu, Yanuca Island, Fiji, August 26–30, 2019, Proceedings, Part I*. 3–15.

[17] Pascal Leroy, Jonathan Pisane, and Damien Ernst. 2022. Value-based CTDE Methods in Symmetric Two-team Markov Game: from Cooperation to Team Competition. In *Deep Reinforcement Learning Workshop NeurIPS 2022*. https://openreview.net/forum?id=wDLQXjkzrS7

[18] Q. Vera Liao and Kush R. Varshney. 2021. Human-Centered Explainable AI (XAI): From Algorithms to User Experiences. *CoRR* abs/2110.10790 (2021). arXiv:2110.10790

[19] Tania Lombrozo. 2007. Simplicity and probability in causal explanation. *Cognitive psychology* 55, 3 (2007), 232–257.

[20] Scott M Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. *Advances in neural information processing systems* 30 (2017).

[21] Dang Minh, H Xiang Wang, Y Fen Li, and Tan N Nguyen. 2022. Explainable artificial intelligence: a comprehensive review. *Artificial Intelligence Review* (2022), 1–66.

[22] Afshin Oroojlooy and Davood Hajinezhad. 2022. A review of cooperative multi-agent deep reinforcement learning. *Applied Intelligence* (2022), 1–46.

[23] Aske Plaat, Walter Kosters, and Mike Preuss. 2023. High-accuracy model-based reinforcement learning, a survey. *Artificial Intelligence Review* (2023), 1–33.

[24] Erika Puiutta and Eric Veith. 2020. Explainable Reinforcement Learning: A Survey. In *4th International Cross-Domain Conference for Machine Learning and Knowledge Extraction (CD-MAKE)*. Springer International Publishing, 77–95.

[25] Tabish Rashid, Mikayel Samvelyan, Christian Schroeder, Gregory Farquhar, Jakob Foerster, and Shimon Whiteson. 2018. QMIX: Monotonic value function factorisation for deep multi-agent reinforcement learning. In *International conference on machine learning*. PMLR, 4295–4304.

[26] Mikayel Samvelyan, Tabish Rashid, Christian Schröder de Witt, Gregory Farquhar, Nantas Nardelli, Tim G. J. Rudner, Chia-Man Hung, Philip H. S. Torr, Jakob N. Foerster, and Shimon Whiteson. 2019. The StarCraft Multi-Agent Challenge. International Foundation for Autonomous Agents and Multiagent Systems, 2186–2188. http://dl.acm.org/citation.cfm?id=3332052

[27] Alexander Sieusahai and Matthew Guzdial. 2021. Explaining deep reinforcement learning agents in the atari domain through a surrogate model. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, Vol. 17. 82–90.

[28] Tom Silver, Ashay Athalye, Joshua B Tenenbaum, Tomás Lozano-Pérez, and Leslie Pack Kaelbling. 2022. Learning Neuro-Symbolic Skills for Bilevel Planning. In *6th Annual Conference on Robot Learning*. https://openreview.net/forum?id=OIaJRUo5UXy

[29] Kyunghwan Son, Daewoo Kim, Wan Ju Kang, David Earl Hostallero, and Yung Yi. 2019. Qtran: Learning to factorize with transformation for cooperative multi-agent reinforcement learning. In *International conference on machine learning*. PMLR, 5887–5896.

[30] Budhitama Subagdja and Ah-Hwee Tan. 2015. Neural modeling of sequential inferences and learning over episodic memory. *Neurocomputing* 161 (2015), 229–242.

[31] Peter Sunehag, Guy Lever, Audrunas Gruslys, Wojciech Marian Czarnecki, Vinicius Zambaldi, Max Jaderberg, Marc Lanctot, Nicolas Sonnerat, Joel Z Leibo, Karl Tuyls, et al. 2018. Value-Decomposition Networks For Cooperative Multi-Agent Learning Based On Team Reward. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*. 2085–2087.

[32] Ah-Hwee Tan, Budhitama Subagdja, Di Wang, and Lei Meng. 2019. Self-organizing neural networks for universal learning and multimodal memory encoding. *Neural Networks* 120 (2019), 58–73.

[33] Philipp Theumer, Florian Edenhofner, Roland Zimmermann, and Alexander Zipfel. 2022. Explainable Deep Reinforcement Learning for Production Control. In *Proceedings of the Conference on Production Systems and Logistics: CPSL 2022*. Hannover: publish-Ing., 809–818.

[34] Pulkit Verma, Shashank Rao Marpally, and Siddharth Srivastava. 2022. Discovering User-Interpretable Capabilities of Black-Box Planning Agents. In *Proceedings of the 19th International Conference on Principles of Knowledge Representation and Reasoning*. 362–372. https://doi.org/10.24963/kr.2022/36

[35] Oriol Vinyals, Timo Ewalds, Sergey Bartunov, Petko Georgiev, Alexander Sasha Vezhnevets, Michelle Yeo, Alireza Makhzani, Heinrich Küttler, John P. Agapiou, Julian Schrittwieser, John Quan, Stephen Gaffney, Stig Petersen, Karen Simonyan, Tom Schaul, Hado van Hasselt, David Silver, Timothy P. Lillicrap, Kevin Calderone, Paul Keet, Anthony Brunasso, David Lawrence, Anders Ekermo, Jacob Repp, and Rodney Tsing. 2017. StarCraft II: A New Challenge for Reinforcement Learning. *CoRR* abs/1708.04782 (2017). arXiv:1708.04782

[36] George A Vouros. 2022. Explainable deep reinforcement learning: state of the art and challenges. *Comput. Surveys* 55, 5 (2022), 1–39.

[37] Wenwen Wang, Budhitama Subagdja, Ah-Hwee Tan, and Janusz A Starzyk. 2012. Neural modeling of episodic memory: Encoding, retrieval, and forgetting. *IEEE transactions on neural networks and learning systems* 23, 10 (2012), 1574–1586.

[38] Xiangjun Wang, Junxiao Song, Penghui Qi, Peng Peng, Zhenkun Tang, Wei Zhang, Weimin Li, Xiongjun Pi, Jujie He, Chao Gao, et al. 2021. SCC: An efficient deep reinforcement learning agent mastering the game of StarCraft II. In *International conference on machine learning*. PMLR, 10905–10915.

[39] Chao Wen, Xinghu Yao, Yuhui Wang, and Xiaoyang Tan. 2020. Smix (λ): Enhancing centralized value functions for cooperative multi-agent reinforcement learning. In *Proceedings of the AAAI Conference on artificial intelligence*, Vol. 34. 7301–7308.

[40] Lulu Zheng, Jiarui Chen, Jianhao Wang, Jiamin He, Yujing Hu, Yingfeng Chen, Changjie Fan, Yang Gao, and Chongjie Zhang. 2021. Episodic multi-agent reinforcement learning with curiosity-driven exploration. *Advances in Neural Information Processing Systems* 34 (2021), 3757–3769.

[41] Weigui Jair Zhou, Budhitama Subagdja, Ah-Hwee Tan, and Darren Wee-Sze Ong. 2021. Hierarchical control of multi-agent reinforcement learning team in real-time strategy (RTS) games. *Expert Systems with Applications* 186 (2021), 115707.