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

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

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

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

ABSTRACT

This paper introduces a method to explain MADRL agents' behaviors by abstracting their actions into high-level strategies. Particularly, a spatio-temporal neural network model is applied to encode the agents' sequences of actions as memory episodes wherein an aggregating memory retrieval can generalize them into a concise abstract representation of collective strategies. To assess the effectiveness of our method, we applied it to explain the actions of QMIX MADRL agents playing a StarCraft Multi-agent Challenge (SMAC) video game. A user study on the perceived explainability of the extracted strategies indicates that our method can provide comprehensible explanations at various levels of granularity.

KEYWORDS

Multi-agent Deep Reinforcement Learning; Explainable Artificial Intelligence; Sequential Decision Making

ACM Reference Format:
Khaing Phyo Wai, Minghong Geng, Shubham Pateria, Budhitama Subagdja, and Ah-Hwee Tan. 2024. Explaining Sequences of Actions in Multi-agent Deep Reinforcement Learning Models: 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

Multi-agent Deep Reinforcement Learning (MADRL) [5, 20, 40] has been demonstrated to solve complex real-world problems such as real-time strategic (RTS) games [3, 8, 22] against human players. However, MADRL models use black-box neural networks which learn massively distributed representations, making the explanation of the learned knowledge challenging [14, 21, 25, 28, 38]. Although various Explainable AI (XAI) [1, 13, 33, 41] methods have been used for interpreting Deep Reinforcement Learning (DRL) [7, 11, 16, 19, 39], existing approaches for explaining MADRL models are still lacking and limited only to offer insights into the agents' cooperative behaviors [9, 37] rather than explaining their coordinated sequences of actions or strategies.

This study aims to explain MADRL models' behavior by interpreting sequences of actions across multiple agents, employing an *explanation by simplification* [17] approach to translate low-level primitive actions into high-level abstract sequences. Specifically, we introduce a spatio-temporal neural network model based on a modified Episodic Memory–Adaptive Resonance Theory (EM-ART) [32, 36] for encoding and generalizing sequences of actions performed by MADRL agents across multiple episodes. We also employ a time-based memory retrieval procedure [4, 10] to generalize encoded actions over time into short abstract sequential patterns, along with a two-stage process for transforming episodes into sequence of significant and unique events.

Empirical evaluation using the StarCraft Multi-Agent Challenge (SMAC) [23] game environment demonstrates that our approach simplifies agent actions into comprehensible strategies. In our previous work [35], we focused on explaining a simple *4t* scenario with four siege tanks in the SMAC environment. In this paper, we extend the task into a more complex *4t8sp* scenario. A comprehensive user study is also included in this paper to assess the perceived explainability of the strategies derived from agents' action sequences.

2 METHODOLOGY

Our proposed framework for explaining opaque MADRL models consists of two main stages, outlined as follows.
Step 1: Memory Encoding. The learned behaviours of the MADRL agents, in terms of sequences of actions performed, are encoded using an episodic memory model, such as EM-ART, which learns the salient action patterns over time.
Step 2: Abstracting the Learned Knowledge. The generalized joint actions and sequences learned in the episodic memory models are extracted and further abstracted into high-level strategies for explanation.

During the memory encoding, traces of actions of the pre-trained MADRL agents are firstly transferred into EM-ART as a memory model to capture and generalize events across space, time, and actions, in the form of episodes (sequences of action events). EM-ART stores events and episodes by combining two fusion ART networks [30, 32]: one for encoding events and the other for episodes [26,

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).

36]. In addition, time stamps of the action events are explicitly encoded using complement coding in a time input field so that an interval-based memory retrieval procedure [4, 10] can be applied to generalize the encoded actions and behaviour patterns of the agents over a selected time interval into abstract sequential patterns. Finally, the abstracted sequences of action events go through a two-stage process in which significant events are selected followed by the removal of repeated events yielding shorter abstract sequences of unique significant events.

3 EXPERIMENTS

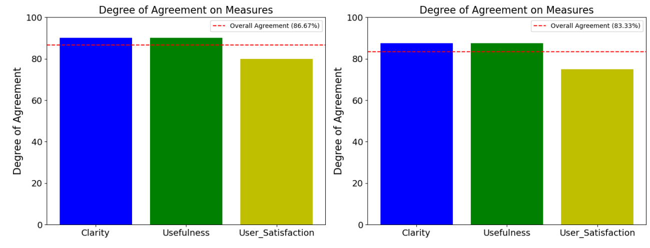
Based on the StarCraft Multi-Agent Challenge (SMAC) [23] platform, we first applied the proposed method to explain gameplays in a scenario named *4t*, wherein four homogeneous siege tanks controlled by MADRL performed combat with four symmetrically positioned enemy units controlled by SC2 AI [2, 12, 15, 34]. We further conducted experiments based on a more complex *4t8sp* scenario, wherein the four tank agents (*4t*) were tasked to overcome the enemy units and reach a predefined target location through eight strategic points (*8sp*).

Table 1: A winning episode for the *4t8sp* scenario derived with event abstraction over two (or more) agents and episode abstraction over time. Legend of actions: N, S, E, and W indicate move north, south, east, and west respectively; A_i indicates attack[enemy_*i*]; and X indicates no_op.

Time Interval	Action	Time Interval	Action
t1-t4	WN	t69-t72	A_0
t5-t8	S	t73-t76	A_3
t9-t16	E	t77-t84	XN
t17-t20	NE	t85-t88	XW
t21-t24	E	t89-t112	XN
t25-t28	N	t113-t132	XE
t29-t32	WN	t133-t140	XN
t33-t36	EN	t141-t148	XE
t37-t48	N	t149-t172	XS
t49-t52	NW	t173-t184	XE
t53-t56	N	t185-t192	XS
t57-t60	A_1	t193-t212	XE
t61-t68	A_2	t213-t254	XN

For the *4t* scenario, we employed QMIX [21] for training the multi-agent teams. For the *4t8sp* scenario, the QMIX agents were further controlled by a class of self-organizing neural networks called Fusion Architecture for Learning and Cognition (FALCON) [29, 31] through the eight strategic points [6]. Based on the actions performed by the QMIX agents after training, we built EM-ART models using different settings of vigilance parameters for event learning and episode learning to study their effects on generalization of events and episodes. We also conducted analysis to identify specific values of the abstraction factor that work best for each scenario.

Table 1 provides an abstracted winning episode for the *4t8sp* scenario, extracted from the EM-ART model using the interval-based memory retrieval algorithm with an abstraction factor of 60. The table illustrates how a sequence of actions taken by the agents over 254 time steps can be condensed into 60 time intervals. This shows that the proposed abstraction method can summarize the complex sequence of the agent actions into a more understandable form, offering a high level perspective on the agent interactions and enhancing accessibility for analysis and interpretation.



(a) *4t* Scenario **(b) *4t8sp* Scenario**

Figure 1: Degree of agreement among participants regarding *clarity, usefulness* and *user satisfaction* that are above the agreement rating threshold (>4).

4 USER STUDY

A user study was conducted by using a method known as *Inter-Rater Agreement Analysis* [18, 24, 27] to examine the impact of explaining action sequences executed by multiple agents in terms of *clarity*, *usefulness*, and *user satisfaction*. The study was conducted via an online survey involving a diverse group of participants varying in age, gender, and familiarity with real-time strategy games. The survey involved the participants reviewing both unexplained and explained gameplay videos and responding to six questions for each of the five distinct games for the SMAC *4t* and *4t8sp* scenarios. Ratings were provided on a Likert scale from 1 (Strongly Disagree) to 5 (Strongly Agree) for assessing the explanation quality.

For the *4t* scenario, the respondents shows a high level of agreement on *clarity*, indicating clear and understandable explanations for the actions taken. The high agreement on *usefulness* suggests that explanations were valuable for understanding actions. Similarly, agreement on *user satisfaction* indicates satisfaction with the provided explanations. Overall, the 86.67% agreement reflects a strong agreement on explanation quality, considering *clarity*, *usefulness*, and *user satisfaction*.

For the *4t8sp* scenario study, the assessment on *clarity* and *usefulness* suggests clear explanations and major consensus on their significance. *User satisfaction*, though lower than both clarity and usefulness, remains reasonably high, indicating overall satisfaction in this more complex scenario. Despite a slightly lower agreement rate compared to the *4t* scenario, the *4t8sp* scenario achieves an overall agreement of 83.33%, signifying substantial agreement among respondents. The results suggest that the explanations were overall well-received and effectively conveyed the sequences of actions by the agents. These findings thus support the effectiveness of the explanation system, even in complex scenarios like *4t8sp*.

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] Sajid Ali, Tamer Abuhmed, Shaker El-Sappagh, Khan Muhammad, Jose M Alonso-Moral, Roberto Confalonieri, Riccardo Guidotti, Javier Del Ser, Natalia Díaz-Rodríguez, and Francisco Herrera. 2023. Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. *Information Fusion* 99 (2023), 101805.

[2] Per-Arne Andersen, Morten Goodwin, and Ole-Christoffer Granmo. 2018. Deep RTS: a game environment for deep reinforcement learning in real-time strategy games. In *2018 IEEE conference on computational intelligence and games (CIG)*. IEEE, 1–8.

[3] Nesma M Ashraf, Reham R Mostafa, Rasha H Sakr, and MZ Rashad. 2021. A state-of-the-art review of deep reinforcement learning techniques for real-time strategy games. *Applications of Artificial Intelligence in Business, Education and Healthcare* (2021), 285–307.

[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] 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.

[6] Minghong Geng, Shubham Pateria, Budhitama Subagdja, and Ah-Hwee Tan. 2023. HiSOMA: A Hierarchical Multi-Agent Model Integrating Self-Organizing Neural Networks with Multi-Agent Deep Reinforcement Learning. *SSRN Electronic Journal* (December 2023). https://dx.doi.org/10.2139/ssrn.4666277

[7] 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.

[8] Isaac Han and Kyung-Joong Kim. 2024. Deep ensemble learning of tactics to control the main force in a real-time strategy game. *Multimedia Tools and Applications* 83, 4 (2024), 12059–12087.

[9] 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.

[10] 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.

[11] Rahul Iyer, Yuezhang Li, Huao Li, Michael Lewis, Ramitha Sundar, and Katia Sycara. 2018. Transparency and explanation in deep reinforcement learning neural networks. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*. ACM, 144–150.

[12] Mingyu Kim, Jihwan Oh, Yongsik Lee, Joonkee Kim, Seonghwan Kim, Song Chong, and Seyoung Yun. 2023. The StarCraft Multi-Agent Exploration Challenges: Learning Multi-Stage Tasks and Environmental Factors Without Precise Reward Functions. *IEEE ACCESS* 11 (2023), 37854–37868.

[13] Sarit Kraus, Amos Azaria, Jelena Fiosina, Maike Greve, Noam Hazon, Lutz Kolbe, Tim-Benjamin Lembcke, Jorg P Muller, Soren Schleibaum, and Mark Vollrath. 2020. AI for explaining decisions in multi-agent environments. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 34. 13534–13538.

[14] Pascal Leroy, Jonathan Pisane, and Damien Ernst. 2022. Value-based CTDE Methods in Symmetric Two-team Markov Game: from Cooperation to Team Competition. *CoRR* abs/2211.11886 (2022). https://doi.org/10.48550/ARXIV.2211.11886 arXiv:2211.11886

[15] Lin Li, Wanzhong Zhao, Chunyan Wang, Abbas Fotouhi, and Xuze Liu. 2024. Nash double Q-based multi-agent deep reinforcement learning for interactive merging strategy in mixed traffic. *Expert Systems with Applications* 237 (2024), 121458.

[16] 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 https://arxiv.org/abs/2110.10790

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

[18] Chu Fei Luo, Rohan Bhambhoria, Samuel Dahan, and Xiaodan Zhu. 2022. Evaluating Explanation Correctness in Legal Decision Making. *Proceedings of the Canadian Conference on Artificial Intelligence* (may 27 2022). https://caiac.pubpub.org/pub/67i6fcki.

[19] Prashan Madumal, Tim Miller, Liz Sonenberg, and Frank Vetere. 2020. Explainable reinforcement learning through a causal lens. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 34. 2493–2500.

[20] Afshin Oroojlooy and Davood Hajinezhad. 2023. A review of cooperative multi-agent deep reinforcement learning. *Applied Intelligence* 53, 11 (2023), 13677–13722.

[21] 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.

[22] Glen Robertson and Ian Watson. 2014. A review of real-time strategy game AI. *Ai Magazine* 35, 4 (2014), 75–104.

[23] 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. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS '19, Montreal, QC, Canada, May 13-17, 2019*, Edith Elkind, Manuela Veloso, Noa Agmon, and Matthew E. Taylor (Eds.). International Foundation for Autonomous Agents and Multiagent Systems, 2186–2188. http://dl.acm.org/citation.cfm?id=3332052

[24] Yael Septon, Tobias Huber, Elisabeth André, and Ofra Amir. 2023. Integrating policy summaries with reward decomposition for explaining reinforcement learning agents. In *International Conference on Practical Applications of Agents and Multi-Agent Systems*. Springer, 320–332.

[25] 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.

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

[27] Budhitama Subagdja, Ah-Hwee Tan, and Yilin Kang. 2019. A coordination framework for multi-agent persuasion and adviser systems. *Expert Systems with Applications* 116 (2019), 31–51.

[28] 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.

[29] Ah -Hwee Tan. 2007. Direct Code Access in Self-Organizing Neural Architectures for Reinforcement Learning. In *Proceedings of The Twentieth International Joint Conference on Artificial Intelligence*. 1071–1076.

[30] Ah-Hwee Tan, Gail A Carpenter, and Stephen Grossberg. 2007. Intelligence through interaction: Towards a unified theory for learning. In *International Symposium on Neural Networks*. Springer, 1094–1103.

[31] Ah-Hwee Tan, Ning Lu, and Dan Xiao. 2008. Integrating Temporal Difference Methods and Self-Organizing Neural Networks for Reinforcement Learning with Delayed Evaluative Feedback. *IEEE Transactions on Neural Networks* 9, 2 (2008), 230–244.

[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] Johanna Vielhaben, Sebastian Lapuschkin, Grégoire Montavon, and Wojciech Samek. 2023. Explainable AI for Time Series via Virtual Inspection Layers. *CoRR* abs/2303.06365 (2023). https://doi.org/10.48550/ARXIV.2303.06365 arXiv:2303.06365

[34] 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

[35] 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. In *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems*. 2325–2327.

[36] 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.

[37] Xinzhi Wang, Huao Li, Rui Liu, Hui Zhang, Michael Lewis, and Katia P. Sycara. 2020. Explanation of Reinforcement Learning Model in Dynamic Multi-Agent System. *CoRR* abs/2008.01508 (2020). arXiv:2008.01508

[38] 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.

[39] Herman Yau, Chris Russell, and Simon Hadfield. 2020. What did you think would happen? explaining agent behaviour through intended outcomes. *Advances in Neural Information Processing Systems* 33 (2020), 18375–18386.

[40] Qiyue Yin, Tongtong Yu, Shengqi Shen, Jun Yang, Meijing Zhao, Kaiqi Huang, Bin Liang, and Liang Wang. 2022. Distributed Deep Reinforcement Learning: A Survey and A Multi-Player Multi-Agent Learning Toolbox. *CoRR* abs/2212.00253 (2022). https://doi.org/10.48550/ARXIV.2212.00253 arXiv:2212.00253

[41] Lingxiang Yun, Di Wang, and Lin Li. 2023. Explainable multi-agent deep reinforcement learning for real-time demand response towards sustainable manufacturing. *Applied Energy* 347 (2023), 121324.