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

SMU-NCC

Khaing Phyo Wai, Minghong Geng, Shubham Pateria, Budhitama Subagdja, Ah-Hwee Tan

Singapore Management University, Singapore

AAMAS 2024 AUCKLAND NZ

Background

- 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; Explainable Deep Reinforcement Learning.

Approach and Methodology

- Our proposed framework for explaining opaque MADRL models consists of two main stages, outlined as follows.

Step 1: Memory Encoding. The learned behaviors of the MADRL agents, in terms of sequences of actions performed, are encoded using an episodic memory model, such as EM-ART [1, 2, 3], 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.

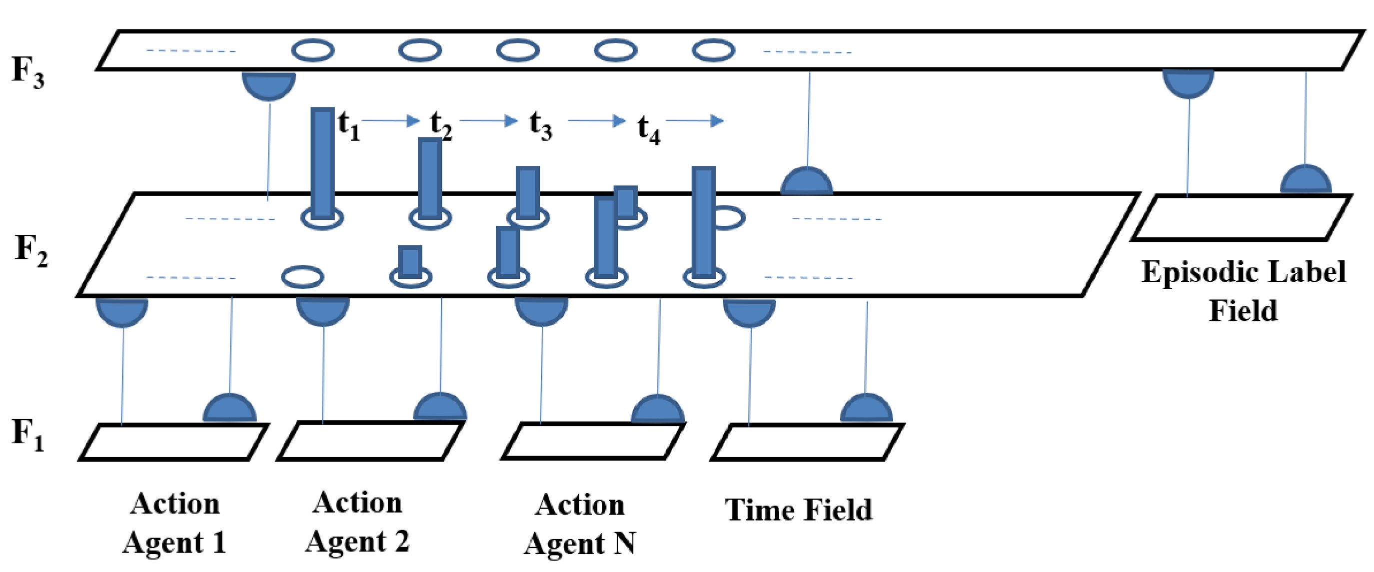


Figure 1: The network architecture of the EM-ART 2 model for encoding sequences of actions performed by multiple agents over time.

- Action events' timestamps are encoded using complement coding in a time input field. This allows for applying an interval-based memory retrieval procedure to generalize agents' encoded actions and behavior patterns over selected time intervals into abstract sequential patterns.
- Subsequently, the abstracted sequences of action events undergo a two-stage process: selecting significant events and removing repeated events, resulting in shorter sequences of unique significant events.

Experiments

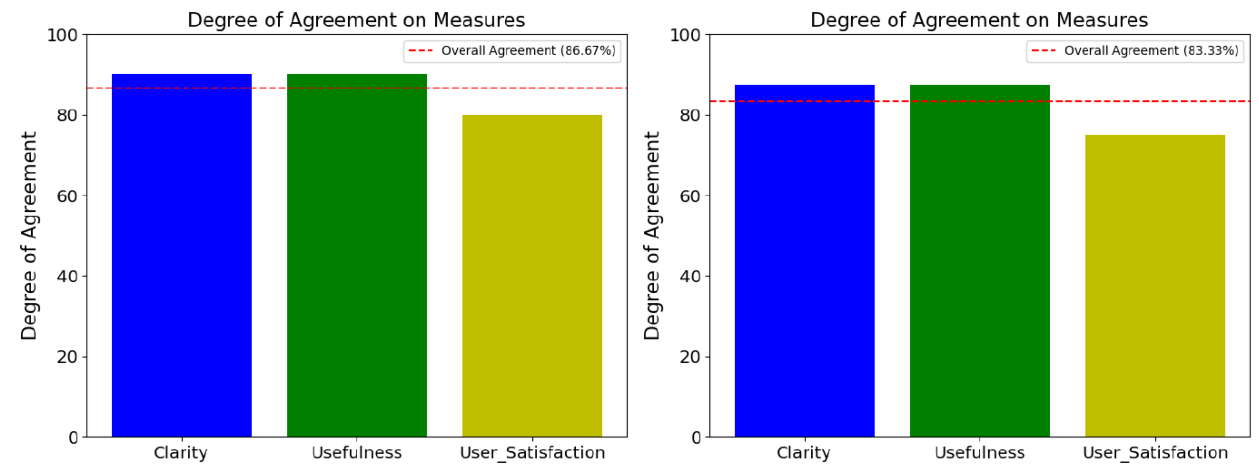
- Using the SMAC [4] platform, the proposed method utilizes QMIX [5] for training and explains gameplay in scenarios ***4t*** and ***4t8sp***, involving four siege tanks against enemy units and strategic point objectives.
- Table 1 presents a winning episode abstraction for the ***4t8sp*** scenario, with an abstraction factor of 60, summarizing agent actions into intervals for enhanced analysis.

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

User Study

- A user study was conducted by using a method known as Inter-Rater Agreement Analysis [6] to examine the impact of explaining action sequences executed by multiple agents in terms of clarity, usefulness, and user satisfaction.
- The study utilized an online survey with a diverse participant group, evaluating unexplained and explained gameplay videos across five distinct games for SMAC's ***4t*** and ***4t8sp*** scenarios, with participants answering six questions per game and providing Likert scale ratings (1 to 5) to assess explanation quality.



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

Figure 2: Degree of agreement among participants regarding clarity, usefulness and user satisfaction that are above the agreement rating threshold (>4).

- Respondents exhibited high levels of agreement on clarity, usefulness, and user satisfaction regarding explanations for actions in the ***4t*** scenario, with an overall agreement of 86.67%.
- In the study of the ***4t8sp*** scenario, assessments revealed clear explanations with significant consensus on their importance, resulting in an overall agreement of 83.33%, slightly lower than the ***4t*** scenario but still indicating substantial consensus 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***.

Conclusion

- This work introduces a methodology using the EM-ART 2 model to interpret sequences of actions by decentralized agents trained with the QMIX model.
- The methodology generalizes and abstracts action sequences using time-based retrieval and two episode abstraction algorithms, effectively explaining multi-agent team behavior.
- It is applicable not only to QMIX-trained MADRL agents but also adaptable to models trained by different algorithms, potentially expediting the learning process.
- Future research aims to incorporate contextual information for deeper insights into MADRL decision-making processes.

Acknowledgement

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] 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.

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

[3] 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.

[4] 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.

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

[6] Joni O Salminen, Hind A Al-Merekhi, Partha Dey, and Bernard J Jansen. 2018. Inter-rater agreement for social computing studies. In 2018 fifth international conference on social networks analysis, management and security (snams). IEEE, 80–87.