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

SMU SINGAPORE MANAGEMENT UNIVERSITY

AAMAS 2023 LONDON

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

Neural and Cognitive Computing Group
Singapore Management University

SMU-NCC

Background

- Multi-Agent Deep Reinforcement Learning (*MADRL*) is effective but difficult to interpret or explain.
- Enplaning *MADRL* model decisions enhances transparency, which is essential to foster trust.
- This paper proposes a method to explain *MADRL* models by utilizing a spatio-temporal neural network model for generalizing action events into abstract strategies.
- Two abstraction algorithms are introduced to further abstract action events across agents and episodes over time.
- The proposed method is evaluated on the StarCraft Multi Agent Challenge (*SMAC*) benchmark task, demonstrating its ability to provide high-level explanations of *MADRL* models across various levels of granularity.

Approach and Methodology

- This study aims to explain the behavior of *MADRL* models through interpreting the sequences of action events performed across multiple agents over time.

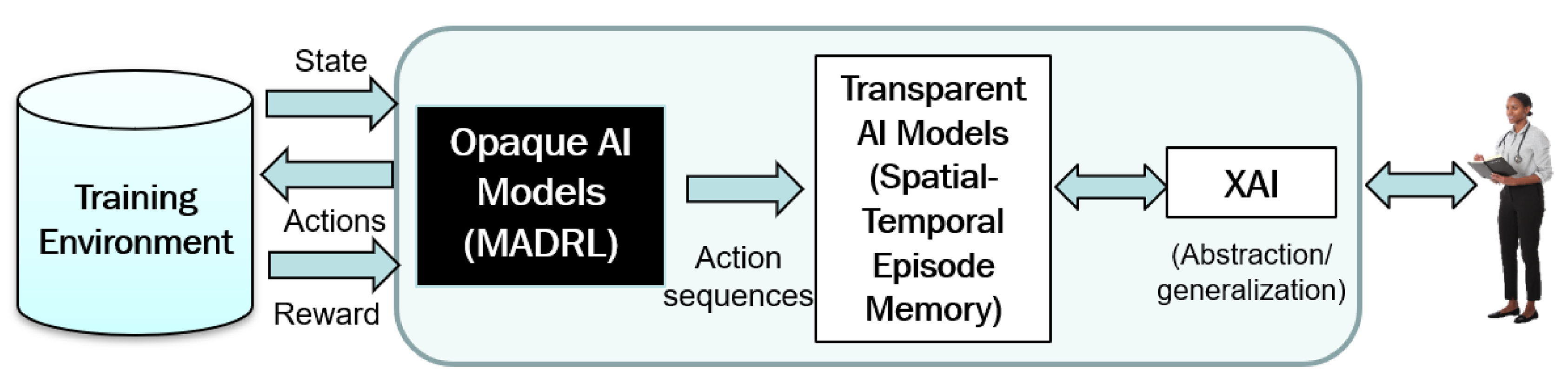

Figure 1: A general framework for explaining opaque *MADRL* models using transparent self-organizing neural network models and post-hoc abstraction algorithms.

- As depicted in Figure 1, our proposed framework for explaining opaque *MADRL* models consists of three main stages, outlined as follows.
  1. **Multi-Agent Deep Reinforcement Learning:** The first step involves training of *MADRL* method.
  2. **Memory Consolidation:** The learned behaviour, namely the sequences of actions performed by the *MADRL* models are encoded using self-organizing neural network models, such as Episodic Memory – Adaptive Resonance Theory (*EM-ART*) model [1, 2] and Spatial-Temporal Episodic Memory (*STEM*) [3] model as illustrated in Figure 2, which learn the salient action patterns and features of their learned behaviors.
  3. **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.

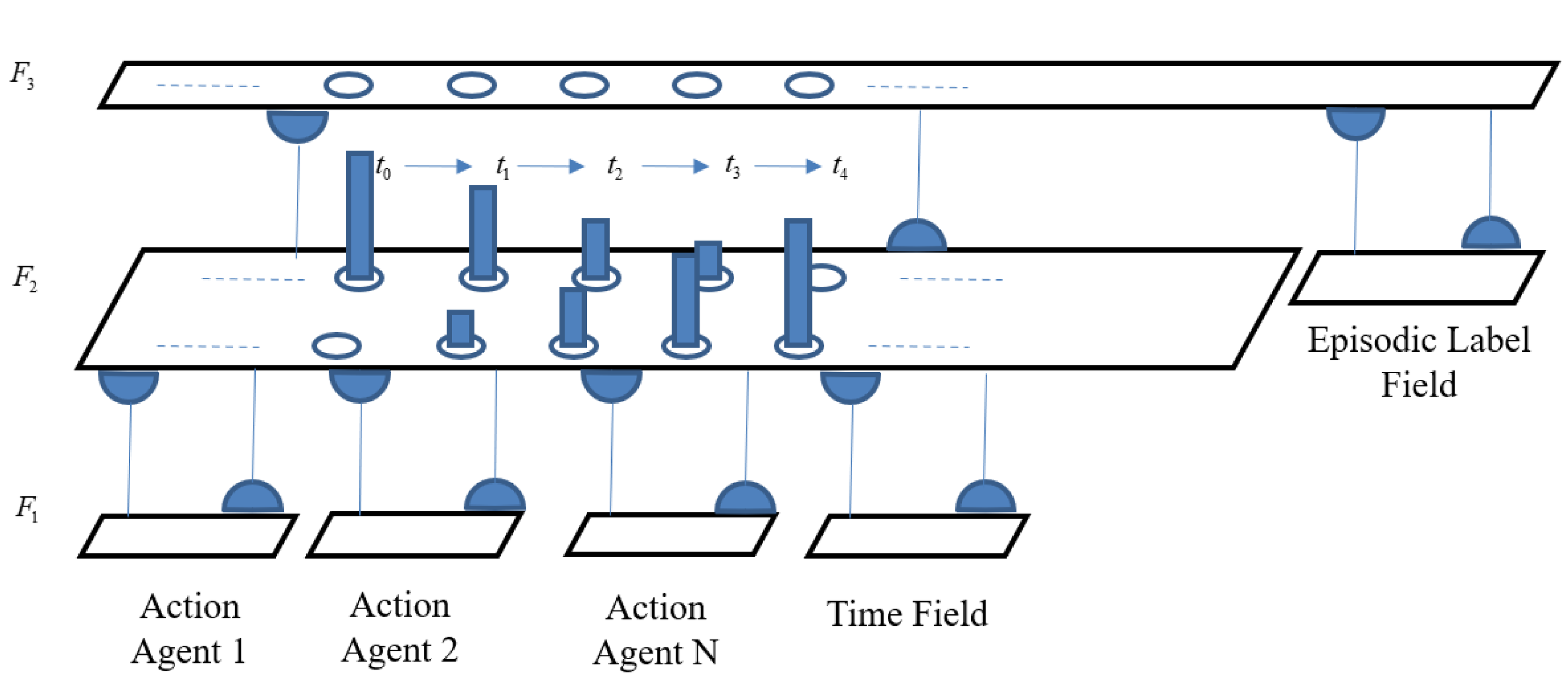

Figure 2: The network architecture of the proposed memory model for encoding sequences of actions performed by multiple agents over time.

- 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).
- The proposed method can provide the explanation of the low-level actions performed by *MADRL* agents in terms of high-level abstracted strategies.

Experiments

- The experiments are conducted using the StarCraft Multi Agent Challenge (*SMAC)*, a set of mini-games developed based on the StarCraft II (*SC2*) real-time strategy (*RTS*) game platform.
- The proposed method is evaluated by using the specific scenario, called ***4t***, in which two groups of four identical units combat each other.
- The *MADRL* model trains the allied group, while the enemy group follows built-in rules.
- In the reported experiments, the agents are homogeneous, capable of performing the same type of actions, and trained using the *QMIX* [4] algorithm, known for its competitive performance in *SMAC.*
- A collection of 1,100 episodes performed by the *MADRL* model is recorded after learning is completed for the experiments.

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.

- *EM-ART* models were constructed using various vigilance parameter values to investigate their influence on the learning and generalization of events and episodes, followed by the utilization of two abstraction algorithms based on the generalized episode.
- Table 1 presents a collection of abstracted events from a winning episode extracted by the *EM-ART* model using an interval-based memory retrieval algorithm with an abstraction factor of 10 and Table 2 displays the final episode abstracted over both agents and time.

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.

- The multi-agent team performs a series of move actions before proceeding with the attack actions.
- At the subgroup level, agents perform strategic positioning before engaging in firing and also perform coordinated attacks on the same enemy unit by attacking enemies in a specific order (A_0, A_3, A_1 followed by A_2).
- The method presented in the paper can explain both single agent and multi-agent models and can be scaled according to the number of agents.

Conclusion

- The proposed approach successfully explains the behavior of the multi-agent team through generalized and abstracted episodes.
- The developed methods can be applied to other scenarios by adjusting the abstraction factor and level of abstraction parameters and can handle episodes of different lengths and any number of agents.

References

[1] Wang, W., Subagdja, B., Tan, A.H. and Starzyk, J.A., 2012. Neural modeling of episodic memory: Encoding, retrieval, and forgetting. IEEE transactions on neural networks and learning systems, 23(10), pp.1574-1586.

[2] Subagdja, B. and Tan, A.H., 2015. Neural modeling of sequential inferences and learning over episodic memory. Neurocomputing, 161, pp.229-242.

[3] Chang, P.H. and Tan, A.H., Encoding and Recall of Spatio-Temporal Episodic Memory in Real Time. Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. Main track. Pages 1490-1496.

[4] Rashid, T., Samvelyan, M., de Witt, C.S., Farquhar, G., Foerster, J. and Whiteson, S., 2018. QMIX: Monotonic value function factorization for deep multi-agent reinforcement learning. In International conference on machine learning. PMLR, 4295–4304.