世界モデルを用いた深層マルチエージェント強化学習における囚人のジレンマ環境下での協力の創発

著者名

November 21, 2025

Abstract

本研究では、深層マルチエージェント強化学習 (DMARL) における協力の創発を、囚人のジレンマ (PD) のようなジレンマ構造が明確な環境下で検証する。既往研究では世界モデルを用いた反事実想像により協調行動の獲得が報告されているが、用いられる環境にはゲーム理論的なジレンマ性が弱いものが含まれる。本研究では、Coin Game のような高次元観測を持つ PD 型環境において、世界モデルが協力創発に寄与しうるか、またそのために必要な拡張 (反事実評価・クレジット割当・因果的報酬成形等) は何かを検討する。

1 Introduction

独立に行動する AI エージェントが本質的に互恵的な性質を備えることは、社会に破滅的な影響を及ぼさないための重要条件である。一方で、AI エージェントは基本的に自己利益的であり、報酬最大化を目標として動作する。

このギャップを埋めるための手法は、古くからゲーム理論の枠組みで研究され、Reputation、Image Score、シグナリングなどが提案されてきた。2000 年代初頭の研究では、エージェントは単純な戦略を用い、レプリケータダイナミクス等を通じて進化ゲーム論の文脈で協力の創発が論じられてきた。

近年は、深層学習・強化学習の進展により、深層マルチエージェント強化学習 (DMARL) と協力の創発に関する研究が急速に活発化している。DMARL では、複数の自律エージェントが環境内で相互作用しながら、それぞれの目的を達成する方策を学習する。しかし、最適方策の獲得は次の理由から困難である。

1. **非定常性**：各エージェントの方策が同時に更新されるため、単一エージェントの MDP で想定される定常性仮定が破られる。
2. **計算複雑性**：次状態の予測には他エージェントの行動推論が不可欠であり、エージェント数の増加とともに複雑さが増大する [1]。

これらの問題に対処する一つの方向として、**世界モデル (World Model)** の応用が進んでいる。たとえば Chai らは、DreamerV2 を拡張し、世界モデル上で反事実想像 (Counterfactual Imagination) を行うことで、複数エージェントの協調を促し、効率的な方策の獲得を報告している [2]。

しかし、既往研究で用いられる環境にはゲーム理論的なジレンマ性が弱いものが含まれる。たとえば HalfCheetah のように役割分担 (前脚・後脚) がある設定では、両者が前進を選ぶときの利得 R が、一方のみが前進 (相手は別行動) したときの利得 T より大きい ($R > T$) と考えられ、**囚人のジレンマ (PD)** に見られる $T > R$ の関係が成立しない。したがって、**明確なジレンマ構造 (PD 型)** における協力創発の検証としては不十分である。

本研究では、PD のようにジレンマ構造が明確な環境において、世界モデルが協力の創発に寄与しうるか、またそのために必要な**拡張 (反事実評価・クレジット割当・因果的報酬成形等)** は何かを検討する。環境としては、Coin Game のように構造はシンプルだが、高次元観測 (例：32×32 のマップ) を持ち、PD 的な利得構造を明示的に設計できる設定を採用する [3]。

2 Related Work

3 Preliminaries

3.1 Multi-Agent Reinforcement Learning

3.2 World Models

3.3 Game Theory and Prisoner's Dilemma

4 Methodology

4.1 Environment Design

4.2 World Model Architecture

4.3 Counterfactual Reasoning and Credit Assignment

5 Experiments

5.1 Experimental Setup

5.2 Results

5.3 Analysis

6 Discussion

7 Conclusion

References

[1] Wong, Annie, Thomas Bäck, Anna V. Kononova, and Aske Plaat. *Deep Multiagent Reinforcement Learning: Challenges and Directions.* Artificial Intelligence Review 56, no. 6 (2023): 5023–56. https://doi.org/10.1007/s10462-022-10299-x

[2] Chai, Jiajun. *Aligning Credit for Multi-Agent Cooperation via Model-Based Counterfactual Imagination.* New Zealand, 2024.

[3] Foerster, Jakob, Richard Y. Chen, Maruan Al-Shedivat, Shimon Whiteson, Pieter Abbeel, and Igor Mordatch. *Learning with Opponent-Learning Awareness.* 2018.