

科学研究費助成事業 研究成果報告書

令和 6 年 6 月 13 日現在

機関番号：15301

研究種目：若手研究

研究期間：2021～2023

課題番号：21K17807

研究課題名（和文）未知の協調・環境を想定したマルチエージェント強化学習の知識転移

研究課題名（英文）Knowledge Transfer in Multi-Agent Reinforcement Learning for Unknown Cooperative Environments

研究代表者

上野 史 (Uwano, Fumito)

岡山大学・環境生命自然科学学域・助教

研究者番号：30880687

交付決定額（研究期間全体）：（直接経費） 3,600,000円

研究成果の概要（和文）：本研究では、未知の協調および環境に対する効率的な知識利用法とそれを利用したマルチエージェント強化学習の方法論を明らかにし、その効果を実験によって実証した。具体的には、環境情報をニューラルネットワークによって抽出し、それを利用した木構造による関数を知識モジュールとして提案し、木構造の関数の枝葉を繋ぎ換えそしてパラメータを強化学習により最適化することで未知の協調および環境に適応した。なお、本研究の成果は英文ジャーナル1件、国際会議2件、招待講演1件、国内学会4件の発表により報告している。

研究成果の学術的意義や社会的意義

本研究は、構造型メタ知識に基づくマルチエージェント学習という従来の研究領域に対して、要素型メタ知識を扱えるように拡張することから、学術的観点から見ると新たな研究領域を開く位置づけとなり、学術的意義が大きい。更に、この基盤技術が確立すれば、ロボットの学習結果を未知問題に転移可能となることから、複数の災害救助ロボットや宇宙探査機による問題解決が可能となる。また、災害救助ロボットで得た学習結果が宇宙探査機に活用可能となるなど、知識の相互利用が可能となるため、産業的意義そして社会的意義も極めて大きい。

研究成果の概要（英文）：In this study, we clarified an efficient method for utilizing knowledge in unknown cooperative and environmental contexts and developed a methodology for multi-agent reinforcement learning based on this approach. We demonstrated its effectiveness through experiments. Specifically, we proposed a knowledge module that extracts environmental information using neural networks and employs a tree-structured function. By reconfiguring the branches and leaves of the tree-structured function and optimizing the parameters through reinforcement learning, we adapted to unknown cooperation and environmental contexts. The results of this research have been reported in one English journal, two international conferences, one invited lecture, and four domestic academic conferences.

研究分野：機械学習

キーワード：マルチエージェントシステム 強化学習 ニューラルネットワーク 未知環境 知識

1. 研究開始当初の背景

近年、目標達成時に得られる報酬を手掛かりに、人が事前に教示せずに適切な行動を学習する強化学習メカニズムを、複数エージェントの協調行動獲得のために各エージェントに導入したマルチエージェント強化学習の実用化が始まっている。特に、実問題のようにエージェントが学習する状況が多数あり、次の状況が予測困難である複雑な問題では、通常マルチエージェント強化学習では協調行動学習が不可能であるため、階層化によるタスク分割と学習結果の抽象化による知識転移が研究されている。しかし、これら従来手法は協調行動や学習環境が「既知」である（エージェントはそれぞれの状況における適切な行動は知らないが、観測情報から状況をどう分割して、各々の状況でどの行動が実行可能であれば問題を解決できるかは知っている）前提で学習しているという問題がある。つまり、実環境を想定した場合、エージェントが環境のどの状況を認識すべきでどの行動をとるべきか全てを把握してシステムを設計することは不可能であり、協調行動や学習環境は「未知」でないことは保証できない。例えば、カーナビ搭載の全ての車の同時経路最適化を考えると、各車の走行経験を活用して交差点を協調して通過することはできるが、他車が事故を起こした場合など走行経験にない状況では活用できず、未知な状況における適切な協調（適切な経路選択）が求められる。また、環境が未知（交通法規の異なる海外）であるときは、自身の経験を適切に変化させた上での協調が求められるため、更に協調が困難になる。

2. 研究の目的

本研究では、従来法では協調行動の学習が不可能な未知の協調・環境を想定したマルチエージェント強化学習法の実現のため、他の問題の学習結果の汎用的知識化とそれに基づく協調行動学習による未知の協調への適応、並びにその転移による未知の環境への展開を目的とする。具体的には、学習結果を各要素に分割し、それぞれを階層的に抽象化することで知識モジュールを設計後、その複数の知識モジュールから新たな知識モジュールを生み出して学習することでエージェント間協調行動を実現する強化学習メカニズムを考案し、その知識モジュールを転移することで未知環境へ展開し、その有効性を検証する。

3. 研究の方法

本研究は、(1)学習結果のモジュール化、(2)知識モジュールに基づく未知の協調行動学習法の提案、(3)未知環境を想定した知識の再構成法を研究課題として取り組む。具体的には、未知の協調および環境に対する効率的な知識利用法とそれを利用したマルチエージェント強化学習の方法論を明らかにし、その効果をロボットシミュレーション実験によって実証する。具体的には、環境情報をニューラルネットワークによって抽出し、それを利用した木構造による関数を知識モジュールとして提案し、木構造の関数の枝葉を繋ぎ換えそしてパラメータを強化学習により最適化することで本研究の目標を達成する。

4. 研究成果

本研究では、まずニューラルネットワークを利用した入力情報の抽象化に取り組み、それを利用したマルチエージェント強化学習を提案した。具体的には、応募者が従来研究してきた報酬設計の技術を応用して、協調行動が未知である環境における学習方法を提案した。また、検証として図1に示すCoin Gameを取り上げ、学習効率を深層強化学習 A3C

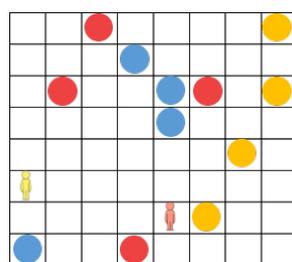


図1 Coin Game

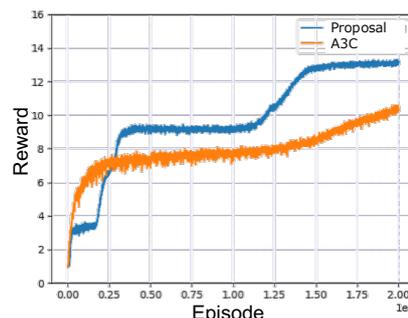


図2 獲得報酬値の推移

(Asynchronous Advantage Actor-Critic) [1]と比較した。Coin Game は各エージェントが自他に限らず同色のコインを取れば正の報酬、それ以外の色のコインを取れば負の報酬が与えられる環境であり、自身と同色のコインのみを取ることが最適な方策となる。図2に結果を示す。結果の通り、提案手法によってより効率的にコインを獲得することが可能となっており、明示的に協調行動が与えられていない未知の協調において、提案手法のニューラルネットワークによる抽象化および学習方法の有効性が示された。

また、未知環境においては、学習結果を効率的に利用可能とするための学習パラメータ調整手法を提案した。具体的には、学習する環境が未知である場合に現在の学習結果を適用させ、利用可能であればそのまま利用し、利用不可能であれば新たな学習を促進させた。実験では2体のエージェントによる迷路問題に適用してその有効性を示した。なお、迷路はある一定の周期でランダムに環境が変化するため、未知環境への適応能力が検証可能となっている。図3にエージェントの獲得した報酬値の合計と全エージェントがゴールへ到達した際にかかるステップ数を示

す。横軸の数値は環境変化のタイミングを表しており、1,000 であれば 1,000 ステップごとに環境変化することを表している。結果から、提案手法が最も精度が高く、未知環境における知識の利用に関してその有効性が示された。

本成果は学習結果のモジュール化を実現する上での基礎となるメタデータの抽出に大きく貢献する結果でありインパクトが大きい。その結果として、本研究成果は第 36 回人工知能学会全国大会、国際会議 EXTRAAMAS 2024 および ICAART 2023 にて発表された。

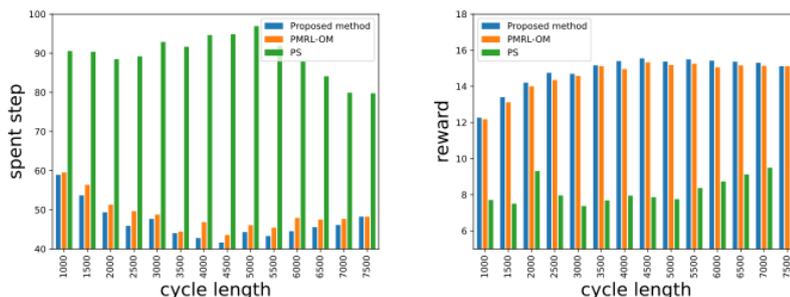


図 3 必要ステップ数と獲得報酬の比較

その後、進化的機械学習 FoRsXCS (Frames-of-References-based XCS) [2]を活用したルールベースの知識表現によって知識モジュールを生成し、知識モジュールの再構成による知識生成法を提案した。具体的には、観測した情報を各要素に抽象化し、それを利用した木構造による関数を知識モジュールとして提案し、木構造の関数の枝葉を繋ぎ換えそしてパラメータを強化学習により最適化することで知識を再構成する。検証には迷路問題におけるロボットのナビゲーションタスクを取り上げた。図 4 は 9 種類の迷路における結果を表しており、緑は提案手法、青は Hi-FoRsXCS[3]、紫は FoRsXCS の結果を示しており、提案手法の方が優れた結果を示している。特に、右図のように Littman 57 という迷路においては頑健な性能を示していることがわかる。以上の成果は、学習結果の知識モジュール化とその再構成による知識生成を可能にしたという点において重要な成果でありインパクトが大きい。その結果として、トップカンファレンスである国際会議 GECCO 2024 にて発表を予定している。

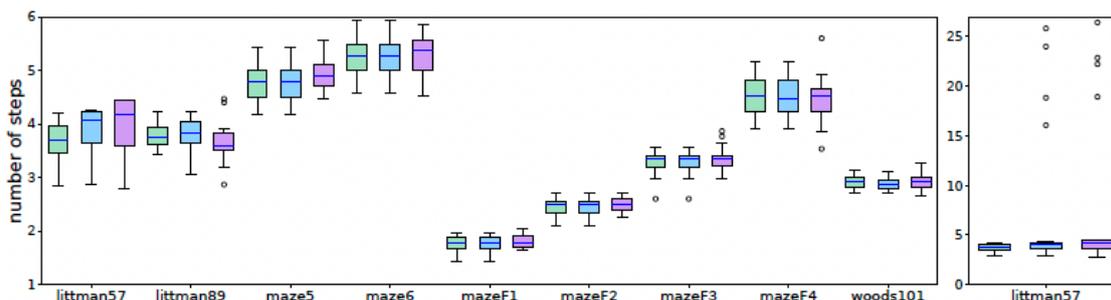


図 4 ロボットナビゲーションタスクにおける結果

以上の研究成果によって、本研究の目標である未知の協調および環境に対する効率的な知識利用法とそれを利用したマルチエージェント強化学習の方法論を明らかにすることができた。今後の課題として、予測による未知の環境および未知の協調への適応が挙げられる。本研究成果では、学習によって未知環境に適応して協調行動を学習したが、ドメインの変更などによって適応不可能となる可能性がある。よって、今後は未知環境・協調を観測情報から推定し、適切に学習可能なカリキュラムを構成することで性能の維持・向上を目指す必要がある。

- [1] Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T. P., Harley, T., Silver, D., and Kavukcuoglu, K., “Asynchronous Methods for Deep Reinforcement Learning,” in *Proceedings of the 33rd International Conference on Machine Learning*, New York, NY, USA, JMLR: W&CP vol. 48, pp. 1928–1937, Jun. 2016.
- [2] Siddique, A., Browne, W. N., and Grimshaw, G. M., “Frames-of-Reference-Based Learning: Overcoming Perceptual Aliasing in Multistep Decision-Making Tasks,” *IEEE Transactions on Evolutionary Computation*, vol. 26, no. 1, pp. 174–187, Feb. 2022.
- [3] Uwano, F. and Browne, W. N., “Hierarchical Frames-of-References in Learning Classifier Systems,” in *Proceedings of the Genetic and Evolutionary Computation Conference Companion 2023*, Lisbon, Portugal, pp. 335–338, Jul. 2023.

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 1件/うち国際共著 0件/うちオープンアクセス 1件）

1. 著者名 上野 史	4. 巻 62
2. 論文標題 観測粒度の違いに適応可能なマルチエージェント強化学習	5. 発行年 2023年
3. 雑誌名 計測と制御	6. 最初と最後の頁 104 ~ 104
掲載論文のDOI（デジタルオブジェクト識別子） 10.11499/sicejl.62.104	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Uwano Fumito, Hasegawa Satoshi, Takadama Keiki	4. 巻 28
2. 論文標題 Inverse Reinforcement Learning with Agents' Biased Exploration Based on Sub-Optimal Sequential Action Data	5. 発行年 2024年
3. 雑誌名 Journal of Advanced Computational Intelligence and Intelligent Informatics	6. 最初と最後の頁 380 ~ 392
掲載論文のDOI（デジタルオブジェクト識別子） 10.20965/jaciii.2024.p0380	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計9件（うち招待講演 1件/うち国際学会 4件）

1. 発表者名 Fumito Uwano
2. 発表標題 Implicit Cooperative Learning on Distribution of Received Reward in Multi-agent System
3. 学会等名 15th International Conference on Agents and Artificial Intelligence (ICAART 2023) (国際学会)
4. 発表年 2023年

1. 発表者名 Fumito Uwano
2. 発表標題 Design of Human-Agent-Group Interaction for Correct Opinion Sharing on Social Media
3. 学会等名 25th International Conference on Human-Computer Interaction (HCI 2022) (国際学会)
4. 発表年 2022年

1. 発表者名 Fumito Uwano
2. 発表標題 Reinforcement Learning in Cyclic Environmental Change for Non-Communicative Agents: A Theoretical Approach
3. 学会等名 5th International Workshop on Explainable and Transparent AI and Multi-Agent Systems (EXTRAAMAS 2023) (国際学会)
4. 発表年 2023年

1. 発表者名 上野 史
2. 発表標題 マルチエージェント強化学習における知識とその境界
3. 学会等名 第69回自律分散システム部会研究会「若手を中心とした模倣学習・強化学習」(招待講演)
4. 発表年 2022年

1. 発表者名 上野 史
2. 発表標題 マルチエージェント強化学習の報酬設計による知識の蒸留と転移に関する一考察
3. 学会等名 第36回人工知能学会全国大会
4. 発表年 2022年

1. 発表者名 上野 史
2. 発表標題 未知の協調・環境を想定したマルチエージェント強化学習の知識転移
3. 学会等名 境界と関係性を視座とするシステムズアプローチ調査研究会
4. 発表年 2021年

1. 発表者名 上野 史
2. 発表標題 獲得報酬の分布に基づくエージェント間の暗黙的協調行動学習とその効果の検証
3. 学会等名 SMASH22 Winter Symposium
4. 発表年 2022年

1. 発表者名 上野 史
2. 発表標題 未知の環境に適応する学習エージェント群の知識利用法の検討
3. 学会等名 計測自動制御学会システム・情報部門学術講演会2021
4. 発表年 2021年

1. 発表者名 Fumito Uwano
2. 発表標題 Cognitive Learning System for Sequential Aliasing Patterns of States in Multistep Decision-Making
3. 学会等名 The Genetic and Evolutionary Computation Conference 2024 (GECCO 2024) (国際学会)
4. 発表年 2024年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------