

令和 6 年 6 月 1 日現在

機関番号：12601

研究種目：挑戦的研究（萌芽）

研究期間：2021～2023

課題番号：21K19816

研究課題名（和文）不完全情報かつ多人数環境で合理的均衡戦略を求める深層強化学習

研究課題名（英文）Deep reinforcement learning in multi-player imperfect-information games

研究代表者

金子 知適（Kaneko, Tomoyuki）

東京大学・大学院総合文化研究科・教授

研究者番号：00345068

交付決定額（研究期間全体）：（直接経費） 4,800,000円

研究成果の概要（和文）：人工知能分野の強化学習では、環境の中で自律的に行動するAIエージェントを想定し、そのエージェントが試行錯誤を通じて振る舞いを学習する技術を扱う。本研究では、不完全情報かつ多人数のゲームを題材に、潜在的な環境のモデルを含めて学習する技術を開発し、広い分野でAIエージェントの強化学習技術の適用を容易にするために研究を行った。数理的なモデルの検討と計算機実験による評価を行い、詳細は、学術論文として発表している。

研究成果の学術的意義や社会的意義

強化学習は、最近ではLLMsのような自然言語を操る巨大AIモデルの学習まで含む、広い応用範囲を持つこれからの社会の基盤技術である。この強化学習技術を成熟させ、誰もが使えるようになることは自由な研究ひいては社会活動のために重要と考えられる。AlphaZeroのような2人完全情報ゲームを対象を限定した場合と比べて、現実の諸問題への強化学習技術の応用はさまざまな難しさを含むが、本研究課題はその解消に向けて2人や完全情報という制限を外した際の難しさを扱った。

研究成果の概要（英文）：This study focuses on reinforcement learning in multi-agent and imperfect information games where AI agents learn a given game throughout trials and errors in playing without human assistance. Our contribution includes mathematical models of agents understanding an environment in latent spaces, computational experiments for performance evaluation, and published papers.

研究分野：ゲームプログラミング

キーワード：ゲームプログラミング

## 様式 C - 19、F - 19 - 1 (共通)

### 1. 研究開始当初の背景

ゲーム研究は人工知能 (AI) の到達度を測る試金石として、古くからチェスや囲碁が注目されてきた。申請者も、これまでそのような二人完全情報ゲームについて研究を行ってきたが、一方で、現実の諸問題を二人完全情報と見立てられる状況は少ない。三人以上の関係者、観測できない情報、不透明な「ルール」のなかでも AI は行動する必要がある。すなわち、本研究で扱う不完全情報かつ多人数ゲームでの強化学習は、ゲームから現実への応用のための自然な一歩である。高性能な AI エージェントを作成するための要素技術には、深層学習と推論 (探索) の二点が重要である。人間トップを超えて有名になった AlphaZero も、ポーカーの Pluribus も、深層学習に基づく直観相当の判断力と、現在の状況の先を検討する能力の両方を備えている。また、深層学習の応用領域では、これまで end-to-end、すなわち問題の入出力のみに着目し問題の内部構造には立ち入らない方式が主流であったが、問題に適した構造を統合したほうが性能が良くなるという報告も増えつつあった。環境の構造を学習する研究が AlphaZero と合流し、MuZero という実験的な研究も生まれていた。

### 2. 研究の目的

人工知能分野の強化学習では、環境の中で自律的に行動する AI エージェントを想定し、そのエージェントが試行錯誤を通じて振る舞いを学習する技術を研究する。本研究の目的は、不完全情報かつ多人数のゲームを題材に、潜在的な環境のモデルを含む深層強化学習の基盤技術を開発し、広い分野で AI エージェントの強化学習技術の適用を容易にすることである。強化学習のアイデアは、汎用性の高い枠組みでエージェントと環境の相互作用を扱う。エージェントが「行動」することで「環境」に働きかけ、環境はエージェントの行動と他の要因により「状態」を確率的に変え、エージェントは状態の一部を観測し、ときおり報酬 (ペナルティを含む) を得る。ここで、どのような状態と行動で報酬が得られるかは既知でないため、エージェントは試行錯誤を通じて環境に適した行動を学ぶ必要がある。強化学習は実際に、ロボット制御から、サッカー、囲碁まで幅広く応用され成果が得られつつある。本研究ではその対象をさらに広げて、現実に近い複雑さを持つ問題の例として、不完全情報かつ多人数のゲームを扱う。不完全情報とは観測できない状態が存在することであり、多人数とは敵とも味方とも想定される他者が複数存在することである。

### 3. 研究の方法

理論的な数理的なモデルの検討と、それを踏まえた具体的なソフトウェアへの導入、計算機実験による性能評価により科学に貢献する。技術としては、多人数環境の特性としてチームメイトとの協調や敵チームとの対峙、不完全情報ゲーム (部分観測問題) の特性として状況の把握や相手に推測されないことなどを考慮に含めたモデルとする必要がある。

#### 4 . 研究成果

重要な戦略空間の効率的な探索、多様性の獲得と活用、各エージェントが独立に判断・行動する中でのチーム全体の協調行動の学習などに関する技術を開発した。具体的には、未知の状態の様々な可能性について、部分的な情報を得るごとにベイズの事後分布として予想を更新し戦略に反映させること、さらにその効率的な表現を用いてさまざまな戦略の候補を効率的に学習すること、局所的な情報とチームメイトとの協調行動を全体の協調行動に反映させること、現在実現している多様性を評価し未知かつ有望と思われる方向を効率的に探索することなどがあげられる。それらの技術は、いくつかのポーカー（テキサスホールデム）の亜種、スタークラフトを元にしたスタークラフトマルチエージェントチャレンジ、ガイスターなどを題材に、多人数の環境としての側面、不完全情報ゲームとしての側面、ならびに両方を持つ環境での性質を評価した。以上の成果は、学術論文として発表している。

## 5. 主な発表論文等

〔雑誌論文〕 計10件（うち査読付論文 10件 / うち国際共著 0件 / うちオープンアクセス 8件）

1. 著者名 Wan and Tang and Tian and Kaneko	4. 巻 -
2. 論文標題 DEIR: Efficient and robust exploration through discriminative-model-based episodic intrinsic rewards	5. 発行年 2023年
3. 雑誌名 IJCAI	6. 最初と最後の頁 4289-4298
掲載論文のDOI (デジタルオブジェクト識別子) 10.24963/ijcai.2023/477	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 -
1. 著者名 Xu and Kaneko	4. 巻 -
2. 論文標題 Curiosity-driven exploration for cooperative multi-agent reinforcement learning	5. 発行年 2023年
3. 雑誌名 IEEE ijcn	6. 最初と最後の頁 1-8
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/IJCNN54540.2023.10191336	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 -
1. 著者名 Chen and Kaneko	4. 巻 -
2. 論文標題 Learning strategies for imperfect information board games using depth-limited counterfactual regret minimization and belief state	5. 発行年 2022年
3. 雑誌名 IEEE international conference on games	6. 最初と最後の頁 486-493
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/CoG51982.2022.9893713	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 -
1. 著者名 Yi and Kaneko	4. 巻 -
2. 論文標題 Improving counterfactual regret minimization agents training in card game cheat using ordered abstraction	5. 発行年 2022年
3. 雑誌名 Advances in computer games 2021	6. 最初と最後の頁 3-13
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-3-031-11488-5_1	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 -

1. 著者名 F. Xu and T. Kaneko	4. 巻 -
2. 論文標題 Local coordination in multi-agent reinforcement learning	5. 発行年 2021年
3. 雑誌名 International conference on technologies and applications of artificial intelligence	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/TAAI54685.2021.00036	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 -

1. 著者名 C. Yi and T. Kaneko	4. 巻 -
2. 論文標題 Improve counterfactual regret minimization agents training by setting limitations of numbers of steps in games	5. 発行年 2021年
3. 雑誌名 26th game programming workshop	6. 最初と最後の頁 117-123
掲載論文のDOI (デジタルオブジェクト識別子) なし	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 -

1. 著者名 Y. Sun and T. Kaneko	4. 巻 -
2. 論文標題 Prediction of werewolf players by sentiment analysis of game dialogue in japanese	5. 発行年 2021年
3. 雑誌名 26th game programming workshop	6. 最初と最後の頁 186-191
掲載論文のDOI (デジタルオブジェクト識別子) なし	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 -

1. 著者名 中屋敷 金子	4. 巻 -
2. 論文標題 ついたて王手どうぶつしょうぎの提案とCFRによる戦略の学習	5. 発行年 2021年
3. 雑誌名 第26回ゲームプログラミングワークショップ	6. 最初と最後の頁 34-41
掲載論文のDOI (デジタルオブジェクト識別子) なし	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0件

〔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
---------	---------