

令和 4 年 5 月 27 日現在

機関番号：12601

研究種目：挑戦的研究（萌芽）

研究期間：2018～2021

課題番号：18K19832

研究課題名（和文）不完全情報かつ多人数参加環境に適した構造を持つ深層強化学習手法の開発

研究課題名（英文）deep reinforcement learning for imperfect and multi-player environments

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・准教授

研究者番号：00345068

交付決定額（研究期間全体）：（直接経費） 4,500,000円

研究成果の概要（和文）：不完全情報かつ多人数のゲームを題材に、モデルを持つ深層強化学習に関する研究を行った。強化学習はAlphaGoの成功で有名なように囲碁やビデオゲームで顕著な成果をあげているが、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。問題が複雑になるほど、エージェントの学習は困難になる。そこで本研究では、既存技術である深層学習に加えて、不完全情報かつ多人数を扱うことに適したモデルの獲得と精密化を行う学習フレームワークを研究した。

研究成果の学術的意義や社会的意義

より広くコンピュータを社会に役立てるために、学習するAIエージェントの研究を行った。一般に、人が完璧なプログラムを事前に準備することはほぼ不可能であるため、コンピュータプログラムあるいはAIエージェント自身が適切な振る舞いを身につけることが望ましい。適切に学習させるためには様々な技術的な困難があり、人の教育でも様々な教育方法とトレードオフがあるように、目的に応じて適切な技術を使い分けたり、新たに開発する必要がある。ここでは、複雑な振る舞いが必要とされる状況の題材として、不完全情報とマルチエージェントゲームを題材として、学習方法を研究した。

研究成果の概要（英文）：This study extends deep reinforcement learning techniques into imperfect and multi-player games. As AlphaZero demonstrated, AI agents can learn decent strategies in perfect information games by reinforcement learning techniques. However, the learning is still difficult in more complicated situations, e.g., where some information is hidden and/or more than two agents are involved. Therefore, this study focused on learning in imperfect and multi-player games as a representative of challenging tasks.

研究分野：ゲーム情報学

キーワード：ゲームプログラミング

## 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、申請者もこれまでに主として、囲碁や将棋など二人完全情報ゲームについて研究を行ってきた。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。つまり、三人以上の人物が関係したり、観測できない情報も多い。さらには「ルール」すらも事前に分かることが多い。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。具体的な対象ドメインで高性能な 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。つい最近に人間トップを超えた AlphaGo Zero (Silver et al., 2017) も、ポーカーの DeepStack (Moravčík et al., 2017) も、深層学習に基づく直観相当の判断力と、直観で得た判断を先を読むことでより正確に改訂する推論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る。たとえば Predictron (Silver et al., 2017) や I2A (imagination augmented agent) では、シンプルな未来予測を深層学習に組み合わせ、ビリヤードや、パズルで成果をあげている。これらを踏まえて本課題では、汎用的な強化学習の枠組みに、深層学習と推論を高い質で統合することを目標とした。強化学習ではゲームのルールは事前には分からないため、推論を行うためには、試行錯誤で環境の構造 (モデル) を推定する必要がある。その部分では、Information Set など不完全かつ多人数の環境で必要となる概念に適した学習モデルに組み込むことで、大幅な性能の改善を実現する。下記の表 1 にまとめたように、本申請課題は各分野で人類を超えた AI と比べると汎用性を志向し、モデルをもつ強化学習という観点では既存研究よりも複雑な対象を扱う。

## 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適切な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、モデルを持つ深層強化学習に関する基盤技術を開発することで、より広い分野において AI エージェントの性能向上を実現させることである。強化学習は、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が未来の総合的な報酬につながるは簡単には分からないため、エージェントは試行錯誤を通じて環境を理解する必要がある。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、状況によって敵にも味方にもなりうる他者が存在することである。

## 3. 研究の方法

問題が複雑になるほど、エージェントの学習は困難になる。そこで本研究では、既存技術である深層学習に加えて、不完全情報かつ多人数を扱うことに適したモデルの獲得と精密化を行う学習フレームワークを、新たに提唱し、核となる技術を確認する。深層学習は高性能なニューラルネットワークを生み出す技術で、画像認識など多くの分野で成果をあげてきている。強化学習においても、深層学習は、ビデオゲームの制御のように反射的な反応の改善に大きな成果をあげた。しかし一般には、AlphaGo で局面の勝率を直感的に判断する深層学習に加えて、モンテカルロ木探索 (MCTS) という囲碁のルールに基づいた探索を補ったように、環境に応じた推

論が必須である。そして申請課題を含む一般的な強化学習の枠組みでは、エージェントにゲームのルールは分かっていないため、推論を補う機構そのものを学習で獲得する必要がある。そこでエージェントの内部に、環境に対応するモデルの学習機構を統合して、モデルの獲得と同時にモデルに基づく長期的な計画の立案技術を研究する。

研究方法は、(1) 理論的な手法の考案と (2) 様々なモデルでの計算機実験による性能評価の繰り返しが基本となる。計算機実験のためには、深層学習のための GPU 計算機も用いる。研究において、実験環境の整備と性能評価を効率的に進めるために、OpenAI gym、「カタンの開発者」、「人狼」などのオープンソースプロジェクトを適切に活用する。研究体制は、研究代表者である申請者自身と申請者の研究室所属の学生（申請時点での人数は大学院生 12 名卒研究生 2 名）で研究を行う。また、必要に応じて、学術分野で一流の研究者と、それぞれのゲームの熟達者の助言を得る。

#### 4. 研究成果

研究の主眼である、不完全情報かつマルチプレイヤーのゲームとして、Catan について (Gendre+ 2020) で、人狼について (Wang+ 2018) と (Sun+ 2021) にて、麻雀について (Long+ 2019, 2020) にて発表した。関係するプレイヤーの数が多いとき、敵もいれば味方もいる。味方と協調するための基礎技術として、StarCraft II (Xu+ 2021), サッカー (Hu+ 2021), カードゲーム (Cao+ 2019) の研究を行い発表した。関連して、ゲームあるいは環境毎に専用のエージェントを訓練するのではなく、共通点をモデル化して理解して学習できることが望ましい。そのための技術について、(Wan+ 2018)や (Kanagawa+ 2019) で発表した。強化学習は比較的長い時系列の因果関係を学習可能だが、不完全情報ゲームでナッシュ均衡を求める有力な手法である Counterfactual regret minimization は、時系列の増加にとっても弱い。Pokerのようにすぐに決着がつくゲームでは成果をあげているが、一般のボードゲームのような数十や百を超える手数 of 歴史を扱うことが現実の計算機では難しい。この弱点を克服して強化学習と組み合わせるための技術について、カードゲーム (Gendre+ 2019), (Yi+ 2019, 2020, 2021) ついたて将棋 (中屋敷+ 2021), ガイスター (Chen+ 2019, 2022) で研究を行い発表した。

なお、当初予期していなかったこととしては、モデルベースの強化学習、すなわち推論と深層強化学習の両方の良いところを活用しようという方向の研究が、筆者の予想以上に世界で進んだことがあげられる。DeepMind チームの MuZero (Schrittwieser et al., 2020), Stochastic Muzero (Antonoglou et al., 2022) や Dreamer (Hafner et al., 2021) が例である。とはいえ、不完全情報性を確率的な状態変化として扱う研究は多くあるが、ナッシュ均衡の概念(相手に読まれてつけ込まれることを防ぐ観点)を統合できたものはなく、引き続き研究が必要である。

## 5. 主な発表論文等

〔雑誌論文〕 計23件（うち査読付論文 22件 / うち国際共著 0件 / うちオープンアクセス 11件）

1. 著者名 Gendre and Kaneko	4. 巻 12533
2. 論文標題 Playing catan with cross-dimensional neural network	5. 発行年 2020年
3. 雑誌名 ICONIP	6. 最初と最後の頁 580-592
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-3-030-63833-7_49	査読の有無 有
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 Nakayashiki and Kaneko	4. 巻 N/A
2. 論文標題 Evaluation of loss function for stable policy learning in dobutsu shogi	5. 発行年 2020年
3. 雑誌名 International conference on technologies and applications of artificial intelligence	6. 最初と最後の頁 175-180
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 Gendre and Kaneko	4. 巻 978-4-907626-46-4 C3804
2. 論文標題 Ceramic: A research environment based on the multi-player strategic board game azul	5. 発行年 2020年
3. 雑誌名 25th game programming workshop	6. 最初と最後の頁 155-160
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 Kanagawa and Kaneko	4. 巻 978-4-907626-46-4 C3804
2. 論文標題 Diverse exploration via infomax options	5. 発行年 2020年
3. 雑誌名 Arxiv	6. 最初と最後の頁 1-21
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 無
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 合田 金子	4. 巻 978-4-907626-46-4 C3804
2. 論文標題 離散行動空間における soft actor-critic の評価	5. 発行年 2020年
3. 雑誌名 第25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 175-180
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 978-4-907626-46-4 C3804
2. 論文標題 逆転の余地を考慮した評価関数の設計とどうぶつしょうぎによる評価	5. 発行年 2020年
3. 雑誌名 第25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 22-29
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 Tomoyuki Kaneko and Takenobu Takizawa	4. 巻 11(3)
2. 論文標題 Computer Shogi Tournaments and Techniques	5. 発行年 2019年
3. 雑誌名 IEEE Transactions on Games	6. 最初と最後の頁 267-274
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/TG.2019.2939259	査読の有無 有
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 Yusaku Mandai and Tomoyuki Kaneko	4. 巻 41(2)
2. 論文標題 RankNet for evaluation functions of the game of Go	5. 発行年 2019年
3. 雑誌名 ICGA Journal	6. 最初と最後の頁 78-91
掲載論文のDOI (デジタルオブジェクト識別子) 10.3233/ICG-190108	査読の有無 有
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 Yuji Kanagawa and Tomoyuki Kaneko	4. 巻 19013855
2. 論文標題 Rogue-Gym: A New Challenge for Generalization in Reinforcement Learning	5. 発行年 2019年
3. 雑誌名 IEEE Conference on Games	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/CIG.2019.8848075	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 Hanhua Zhu and Tomoyuki Kaneko	4. 巻 19279615
2. 論文標題 Deep Residual Attention Reinforcement Learning	5. 発行年 2019年
3. 雑誌名 International Conference on Technologies and Applications of Artificial Intelligence	6. 最初と最後の頁 1-6
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/TAAI48200.2019.8959896	査読の有無 有
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 Zhejie Hu and Tomoyuki Kaneko	4. 巻 19279598
2. 論文標題 Application of Deep-RL with Sample-Efficient Method in Mini-games of StarCraft II	5. 発行年 2019年
3. 雑誌名 International Conference on Technologies and Applications of Artificial Intelligence	6. 最初と最後の頁 1-6
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/TAAI48200.2019.8959866	査読の有無 有
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 Chen Chen and Tomoyuki Kaneko	4. 巻 19279608
2. 論文標題 Acquiring Strategies for the Board Game Geister by Regret Minimization	5. 発行年 2019年
3. 雑誌名 International Conference on Technologies and Applications of Artificial Intelligence	6. 最初と最後の頁 1-6
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/TAAI48200.2019.8959878	査読の有無 有
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 Hyunwoo Oh and Tomoyuki Kaneko	4. 巻 -
2. 論文標題 Deep Recurrent Q-Network with Truncated History	5. 発行年 2018年
3. 雑誌名 IEEE Technologies and Applications of Artificial Intelligence	6. 最初と最後の頁 34--39
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/TAAI.2018.00017	査読の有無 有
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 Tianhe Wang and Tomoyuki Kaneko	4. 巻 -
2. 論文標題 Application of Deep Reinforcement Learning in Werewolf Game Agents	5. 発行年 2018年
3. 雑誌名 IEEE Technologies and Applications of Artificial Intelligence	6. 最初と最後の頁 28--33
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/TAAI.2018.00016	査読の有無 有
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 Hanhua Zhu and Tomoyuki Kaneko	4. 巻 -
2. 論文標題 Playing the Flappy Bird with Reinforcement Learning Algorithms	5. 発行年 2018年
3. 雑誌名 The 23rd Game Programming Workshop	6. 最初と最後の頁 153--159
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 Chen Chen and Tomoyuki Kaneko	4. 巻 -
2. 論文標題 Counterfactual Regret Minimization for the Board Game Geister	5. 発行年 2018年
3. 雑誌名 The 23rd Game Programming Workshop	6. 最初と最後の頁 137--144
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 ZheJie Hu and Tomoyuki Kaneko	4. 巻 -
2. 論文標題 Reinforcement Learning with Effective Exploitation of Experiences on Mini-Games of StarCraft II	5. 発行年 2018年
3. 雑誌名 The 23rd Game Programming Workshop	6. 最初と最後の頁 168--174
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 金川裕司 金子 知適	4. 巻 -
2. 論文標題 ローグライクゲームによる強化学習ベンチマーク環境Rogue-Gymの提案	5. 発行年 2018年
3. 雑誌名 第23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 120--127
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 王 天鶴 金子 知適	4. 巻 -
2. 論文標題 人狼エージェントにおける深層Qネットワークの応用	5. 発行年 2018年
3. 雑誌名 第23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 16--22
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 Oh Hyunwoo 金子 知適	4. 巻 -
2. 論文標題 LSTM の初期状態の学習による DRQN の改善	5. 発行年 2018年
3. 雑誌名 第23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 220--227
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 Yi, C. and Kaneko, T.	4. 巻 -
2. 論文標題 Improving counterfactual regret minimization agents training in card game cheat using ordered abstraction	5. 発行年 2021年
3. 雑誌名 Advances in computers and games	6. 最初と最後の頁 印刷中
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 Xu, F. and Kaneko, T.	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 印刷中
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 Hu, Z. and Kaneko, T.	4. 巻 -
2. 論文標題 Hierarchical advantage for reinforcement learning in parameterized action space	5. 発行年 2021年
3. 雑誌名 IEEE international conference on games	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/CoG52621.2021.9619068	査読の有無 有
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 -

〔学会発表〕 計16件 (うち招待講演 0件 / うち国際学会 3件)

1. 発表者名 Yi and Kaneko
2. 発表標題 Improve counterfactual regret minimization for card game cheat
3. 学会等名 25th game programming workshop (国際学会)
4. 発表年 2020年

1. 発表者名 Chen and Kaneko
2. 発表標題 Application of dream to the board game geister
3. 学会等名 25th game programming workshop (国際学会)
4. 発表年 2020年

1. 発表者名 Honghai and Kaneko
2. 発表標題 Training japanese mahjong agent with two dimension feature representation
3. 学会等名 25th game programming workshop (国際学会)
4. 発表年 2020年

1. 発表者名 徐 金子
2. 発表標題 ProcgenBenchmark における汎化性能を高める強化学習
3. 学会等名 第25回ゲームプログラミングワークショップ
4. 発表年 2020年

1. 発表者名 Chen Chen and Tomoyuki Kaneko
2. 発表標題 Utilizing History Information in Acquiring Strategies for Board Game Geister by Deep Counterfactual Regret Minimization
3. 学会等名 The 24th Game Programming Workshop
4. 発表年 2019年

1. 発表者名 Yu Cao and Tomoyuki Kaneko
2. 発表標題 An Extension of Counterfactual Regret Minimization for Multiplayer Card Games
3. 学会等名 The 24th Game Programming Workshop
4. 発表年 2019年

1. 発表者名 Cheng Yi and Tomoyuki Kaneko
2. 発表標題 Performance of Counterfactual Regret Minimization with Self-Confirming Equilibrium
3. 学会等名 The 24th Game Programming Workshop
4. 発表年 2019年

1. 発表者名 中屋敷 太一 金子 知適
2. 発表標題 どうぶつしょうぎを用いた AlphaZero の手法の調査
3. 学会等名 第24回ゲームプログラミングワークショップ
4. 発表年 2019年

1. 発表者名 徐 凡超 金子 知適
2. 発表標題 スタークラフト II のミニゲームにおけるマルチタスク強化学習
3. 学会等名 第24回ゲームプログラミングワークショップ
4. 発表年 2019年

1. 発表者名 ZheJie Hu and Tomoyuki Kaneko
2. 発表標題 Enhancing Sample Efficiency of Deep Reinforcement Learning to Master the Mini-games of StarCraft II
3. 学会等名 The 24th Game Programming Workshop
4. 発表年 2019年

1. 発表者名 Quentin Gendre and Tomoyuki Kaneko
2. 発表標題 Counterfactual Regret Minimisation for playing the multiplayer bluffing dice game Dudo
3. 学会等名 The 24th Game Programming Workshop
4. 発表年 2019年

1. 発表者名 Hanhua Zhu and Tomoyuki Kaneko
2. 発表標題 Training Agents with Long-range Information in Deep Reinforcement Learning
3. 学会等名 The 24th Game Programming Workshop
4. 発表年 2019年

1. 発表者名 Tang Jiachen and Tomoyuki Kaneko
2. 発表標題 Back Prediction in the Game of Go
3. 学会等名 The 24th Game Programming Workshop
4. 発表年 2019年

1. 発表者名 Long Honghai and Tomoyuki Kaneko
2. 発表標題 Improving Mahjong Agent by Predicting Types of Yaku
3. 学会等名 The 24th Game Programming Workshop
4. 発表年 2019年

1. 発表者名 Sun, Y. and Kaneko, T.
2. 発表標題 Prediction of werewolf players by sentiment analysis of game dialogue in japanese
3. 学会等名 The 26th game programming workshop
4. 発表年 2021年

1. 発表者名 中屋敷, 金子
2. 発表標題 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
3. 学会等名 第26回ゲームプログラミングワークショップ
4. 発表年 2021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
---------	---------