様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　６ 年　 ５ 月 ３１ 日現在

機関番号：１２６０１
研究種目：基盤研究(B)（一般）
研究期間：2021～2023
課題番号：２１Ｈ０３５７０

研究課題名（和文）世界モデルの獲得と多様な戦略の探索による深層強化学習の汎用性向上

研究課題名（英文）Onto generalization of reinforcement learning agents through diversity and understanding of 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・教授

研究者番号：００３４５０６８

交付決定額（研究期間全体）：（直接経費）　　5,000,000円

研究成果の概要（和文）：思考ゲームを題材に、AlphaZero のように知識無しの状態から試行錯誤を通じてエージェントに適切な振る舞いを自発的に身につけさせる強化学習の研究を進めた。特に、将棋、囲碁、2048において、理論的な数理モデルの検討と計算機プログラムとしての実装、計算機実験による性能の評価を行った。成果の一部はオープンソースのプログラムと事前学習済みモデルとしてすでに公開済みであり、残りの部分も今後整備を進めて公開する予定である。

研究成果の学術的意義や社会的意義
強化学習は、教師あり学習で必要となる教師データを必要としない代わりに、試行錯誤の経験を積むための計算機資源を必要とする。この研究の遠い目標は、高性能なAIエージェントを少数の巨大組織のみが開発できる状況を変え、個人が自分自身のハードウェアで自分だけのエージェントを持てるようにすることにある。ゲームという限られた対象限定ではあるが、本研究は、一般的なハードウェアで十分に強いエージェントを作成可能となったことに貢献している。

研究成果の概要（英文）：This study focuses on reinforcement learning in perfect information games where AI agents master a given game throughout trials and errors in playing without human assistance such as AlphaZero. Our contribution includes mathematical models, implementation in computer software, and computational experiments for performance evaluation. A part of our results is already available as an open-source software with pre-trained models and more will become available in future.

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

思考ゲーム、特に囲碁、将棋、チェスなどの二人完全情報ゲームでは、ここ数年の AlphaGo から AlphaZero までの研究で、既存手法で作成された AI プレイヤや人間の強さを越えて、大きな技術の進歩があった。一連の研究はもちろん偉大な成果であるが、一方で実用を考えると、この手法を決定版として採用することは難しい。初めに指摘された点は、再現実験に Google (DeepMind)社か同等の巨大企業のみに調達可能な、莫大な計算資源を必要とすることである。Facebook 社、Tencent 社、あるいは世界のボランティアによる分散計算などのいくつかの再現プロジェクトにより幾多の工夫が考案され、また申請者も貢献をしているが、まだ決定的な手法の改善にはいたっていない。もう一つの弱点は、やや後で明らかになり始めたことだが、理論的な基盤が貧弱なことである。AlphaZero で採用されたモンテカルロ木探索と深層強化学習の組み合わせ方について発表当時は自然に受け止められたが、深層強化学習としては数理的により良いモデル化が可能であることが少し後で示されている。理論的な着眼を糸口に、更に深く切り込むことで大幅な性能向上と汎用性の獲得を目指す好機であった。

２．研究の目的

本応募課題では、思考ゲームを題材に、深層強化学習の汎用性向上の研究を行う。同時に、学習 効率の向上を通じて、必要な計算資源の削減を目指す。囲碁、将棋、チェスなどの思考ゲームは AI の到達度を測る試金石として、また AlphaGo が登場して以降は深層強化学習の題材としても 注目を集めている。ロボット制御などでエージェントは知覚に対する反射的な行動を学ぶことに 対し、思考ゲームでは熟慮して判断の質を高めることが重要でそのことを学ぶためである。その ためには、エージェントが確率的な推論や思考実験をできるような世界のモデルが必要となる。 推論に適した世界のモデルの学習には様々な課題があるが、抽象化と多様性を技術的な核に総合的な学習フレームワークを構築する。抽象的な表現で世界を学ぶことは汎用性と学習効率につな がり、環境の多様性とエージェントの多様性が適切な抽象度に誘導するという関係にある。

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。技術的な核を抽象化と多様性におく。直感的には次のように。世界の性質を 抽象的に理解することができれば、各事例を個別に理解するよりも、効率と汎用性の二つの点でメリットがある。抽象的な理解をうるためには、多様な経験が大切である。多様性の具体には、 複数の種類を想定する。主要な一つは環境の多様性で、囲碁の 9 路盤と 19 路盤、将棋とチェスな ど、複数のゲームを同時に学ぶことなどで実現される。また、環境を固定して異なる課題（たとえ ば対局と詰将棋）を行うことも、ゆるやかに環境の多様性を実現する。環境の多様性と対応して、 エージェントの多様性も重要である。一人の人間

には地球の全てを歩いて回れないように、強化 学習の状態空間も広いためエージェントが経験できる範囲も限られる。そのため、エージェント が経験した世界をモデル化するという強化学習のアプローチでは、本質的に世界の一部分しかモ デル化出来ないという制約がある。その前提で実用的な性能を実現するためには、なんらかの方 法で重要な状態を判別してその付近を重点的に経験する必要がある。このために、異なる視点や 興味を持つ、多様なエージェントが有用である。また多様なエージェントは、多数決により集団全体の安定性も実現する。

４．研究成果

AlphaZero とその後継のさまざまな MuZero のプロジェクトが対象としている評価環境の中から、囲碁、将棋、そして 2048 を主に選んで研究成果の適用と性能評価を行った。特に将棋では、数枚の GPU の１週間程度の計算で十分に強い AI エージェントを生み出す、効率の良い強化学習アルゴリズムの開発と実装に成功している。成果は順次、オープンソースプログラムと訓練済みモデルとして公開予定である。

５．主な発表論文等

〔雑誌論文〕　計9件（うち査読付論文　9件／うち国際共著　0件／うちオープンアクセス　4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 ijcnn	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 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 T. Nakayashiki and T. Kaneko	４．巻 -
２．論文標題 Maximum entropy reinforcement learning in two-player perfect information games	５．発行年 2021年
３．雑誌名 IEEE SSCI	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/SSCI50451.2021.9659991	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 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 Z. Hu and T. Kaneko	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 －

1．著者名 H. Zhu and T. Kaneko	4．巻 -
2．論文標題 Residual network for deep reinforcement learning with attention mechanism	5．発行年 2021年
3．雑誌名 J. Inf. Sci. Eng.	6．最初と最後の頁 517-533
掲載論文のDOI（デジタルオブジェクト識別子） 10.6688/JISE.202105_37(3).0002	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 山下 金子	4．巻 -
2．論文標題 4x3 盤面の 2048 の完全解析	5．発行年 2023年
3．雑誌名 第28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 1-5
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 出村 金子	4．巻 -
2．論文標題 将棋における変則ルール「将棋 81 万」の提案と深層強化学習への応用	5．発行年 2023年
3．雑誌名 第28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 111-118
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